



**The Role of Aptitude in Second Language Segmental Learning:
The Case of Japanese Learners' English /ɹ/ Pronunciation
Attainment in Classroom Settings**

Kazuya Saito¹

Abstract

Building on the extensive work conceptualizing, developing, and validating foreign language aptitude for successful second language acquisition (e.g., Skehan, 2015 in *Applied Linguistics*), the current project expounded the cognitive abilities relevant and instrumental to adult Japanese classroom learners' pronunciation attainment of English /ɹ/. The speech production of 50 Japanese participants was elicited, acoustically analyzed, and linked to different aspects of their aptitude profiles (phonemic coding, associative memory, sequence recognition). Whereas those with higher phonemic coding demonstrated better performance in a relatively easy dimension of English /ɹ/ pronunciation (tongue retraction), those with greater associative memory demonstrated more advanced performance in the relatively difficult dimensions of English /ɹ/ pronunciation (longer transition duration, labial/alveolar/pharyngeal constrictions). The role of incidental learning aptitude—sequence recognition—remained unclear. The findings here indicate that explicit aptitude related to phonological analysis and memory may play a key role in predicting the incidence of advanced L2 segmental proficiency attainment in classroom settings.

Key words. Aptitude, foreign language learning, second language speech, pronunciation, English /ɹ/

¹ The current study was partially funded by the Grant-in-Aid for Scientific Research in Japan and Birkbeck College Additional Research Support Grant. I am grateful to the journal editor, Rachel Hayes-Harb, and *Applied Psycholinguistics* reviewers for providing constructive comments on earlier versions of this paper.

To date, scholars have extensively examined what kinds of learners are cognitively adept at learning second language (L2) morphosyntax. However, surprisingly little is known about the cognitive correlates of successful L2 pronunciation learning. In the context of 50 Japanese L2 learners in classroom settings, the current study examined the roles of three different kinds of foreign language aptitude (phonemic coding, associative memory, sequence recognition) in three different aspects of their English /ɪ/ pronunciation acquisition (degree and rate of tongue retraction, labial, alveolar and pharyngeal constrictions).

Background

Foreign Language Aptitude

Many theories of second language acquisition (SLA) (e.g., Ellis, 2006 for usage-based accounts) share the fundamental view that adult L2 learners improve their proficiency as a function of increased language exposure and practice. At the same time, there is a great amount of empirical evidence that the final outcome of L2 proficiency after years of learning is subject to a great deal of individual variability; this is the case in English-as-a-Foreign Language (EFL) classrooms, where access to the target language is limited to just a few hours of instruction a week; and where learners are left without opportunities to meaningfully converse with other native and non-native speakers (Muñoz, 2014). In order to understand the underlying reasons for this variability, scholars have extensively examined a range of cognitive and perceptual abilities (i.e., measures of aptitude) that may contribute to determining the extent to which certain L2 learners can continuously enhance their L2 performance and eventually attain relatively advanced proficiency (see Li, 2015a; and Skehan, 2015 for meta-analytic and narrative reviews).

Over the past 50 years, several aptitude frameworks have been conceptualized, tested and refined. For example, Carroll and Sapon (1959) developed the Modern Language Aptitude Test (MLAT) to assess L2 learners' intentional learning abilities for successful classroom L2 learning, including phonemic coding (the ability to associate sound strings with their corresponding symbols) and associative memory (the ability to associate letters with their corresponding objects). According to their validation reports (e.g., Carroll, 1962), it was found that students who obtained higher MLAT scores were likely to show better test performance and obtain higher final grades in their foreign language classes in the US (see also Alderson, Clapham, & Steel, 1997).

Loosely building on the MLAT (Carroll & Sapon, 1959), Meara (2005) developed the LLAMA—the aptitude test format that has been most widely used in SLA studies and adopted in the current study. The LLAMA is designed to measure various dimensions of L2 learners' abilities to learn a new language using linguistic materials adapted from a British-Columbian indigenous language. The LLAMA adopts the same measures as in the MLAT for assessing intentional learning abilities (i.e., phonemic coding, associative memory). These intentional test scores have been found to predict successful L2 morphosyntax learning especially when L2 learners receive explicit instruction/corrective feedback (e.g., Yalçın & Spada, 2016; Yilmaz & Granena, 2015)

Different from the traditional paradigm (e.g., MLAT), however, the LLAMA adopts a new task which is hypothesized to tap into one form of incidental learning ability—i.e.,

sequence recognition (the ability to remember novel linguistic patterns without awareness). L2 learners' sequence recognition performance has been found to demonstrate strong associations with the degree of their learning success in certain contexts, where both intentional and incidental learning processes play a key role. These instances include the final quality of early bilinguals' L2 morphosyntax performance in naturalistic settings (Granena, 2013), and the longitudinal development of classroom L2 learners' oral proficiency when they engaged in both form-oriented and content-based instruction over time (Saito, Suzukida, & Sun, 2018).

Aptitude Effects on Different Stages of L2 Learning

For the purpose of theory-building on the cognitive individual differences in SLA, Skehan (2016) proposed the acquisition-aptitude model, suggesting that different types of aptitude (e.g., phonemic coding, associative memory, sequence recognition) are uniquely tied to different stages of L2 learning: (i) analyzing incoming input → (ii) automatizing partially acquired knowledge → (iii) attaining advanced-level use of the language. As such, Skehan argued that intentional language analysis, memory and incidental learning abilities could differentially impact and relate to L2 learners' performance, while each stage interconnects with each other in a complementary fashion.

In this model, L2 learners with higher auditory processing abilities (e.g., phonemic coding) are assumed to not only hold more information about unfamiliar sounds in their phonological buffer (i.e., short-term memory), but also analyze it in a prompt and timely fashion (i.e., executive control) (Yilmaz & Koylu, 2016). As L2 learners with greater associative memory can maintain and control larger amounts of information, they are assumed to pay attention to not only perceptually salient but also subtle and infrequent features in the L2 input, which is critical in the later stages of L2 learning (Schneiderman & Desmarai, 1988). To attain highly advanced L2 proficiency, however, L2 learners may need not only strong intentional, but also solid incidental learning abilities to make the most of their L2 experience, as learners at this level may need to automatically detect novel sound, word and grammar patterns during their exposure to the target language, regardless of the presence of awareness (e.g., sequence recognition) (Granena, 2013). Skehan's model stressed that the interaction between aptitude and SLA (phonemic coding × input processing, associative memory × automatization, sequence recognition × lexicalization) could take place simultaneously (rather than independently), as all aspects of L2 learners' developing system are essentially interwoven with each other.

A growing number of studies have investigated the intricate connections between different types of aptitude and multiple aspects of L2 morphosyntax learning. In the context of EFL students in Turkey, for example, Yalçın and Spada (2016) found that those with higher language analysis abilities (grammatical inferencing) demonstrated strong gains in the acquisition of a relatively easy structure (past progressive) after four hours of instruction. Comparatively, those with higher memory abilities (associative memory) successfully acquired the relatively difficult structure (passive). Similarly, Li (2013, 2015b) found that language analysis abilities (grammatical sensitivity) were facilitative of L2 Chinese learners' acquisition of a relatively easy structure (classifiers), especially when their instruction did not involve any metalinguistic explanation nor explicit correction. The current study was

designed to conduct exploratory analyses with regard to the aptitude-acquisition interaction in the context of L2 pronunciation development.

Roles of Aptitude in L2 Pronunciation Learning

Whereas much attention has been directed towards examining the role of aptitude in the learning of L2 morphosyntax, some studies have begun to delve into the cognitive individual differences in L2 phonological acquisition. For example, Darcy and her colleagues conducted a series of empirical studies showing that L2 learners' vowel perception and production performance was associated with a range of cognitive abilities, such as working memory, attention control and processing speed (Darcy, Park, & Yang, 2015), and inhibitory control (Darcy, Mora, & Daidone, 2016). In an intervention study, Li and DeKeyser (2017) demonstrated that musical talent (sensitivity to rhythm) mediated the effects of explicit instruction on American learners' acquisition of L2 Mandarin lexical tones.

Importantly, these cognitive abilities are not necessarily specific to language learning behaviours per se, but rather amenable to many different areas of general skill acquisition (for further discussion on domain-specific vs. domain-general cognitive abilities, see Skehan, 2016). As reviewed earlier, aptitude is defined as a set of capacities directly related to intentional and incidental language learning (Carroll & Sapon, 1959; Meara, 2005). In their synthesis of the L2 speech literature, Trofimovich, Kennedy and Foote (2015, p. 354) pointed out that "there has been little systematic research on the relationship between various subcomponents of language aptitude and L2 pronunciation learning." To advance the agendas of L2 speech aptitude research in an interdisciplinary fashion, more studies are needed to elucidate the extent to which the existing aptitude frameworks in SLA are applicable to the attainment of L2 pronunciation proficiency in classroom settings.

Japanese Learners' English /ɪ/ Acquisition

The current study focused on one of the most well-researched topics in the L2 pronunciation learning literature—the acquisition of English /ɪ/ by adult native Japanese speakers. American English /ɪ/ can be acoustically characterized by (a) lower third formant (F3) (1600-1900 Hz), (b) lower second formant (F2) (1300-1600 Hz) and (c) longer transitional duration of first formant (F1) (50-100 ms) (Espy-Wilson, 1993; Espy-Wilson et al., 2000; Flege, Takagi, & Mann, 1995; Hattori & Iverson, 2009). Since English /ɪ/ is absent in the Japanese approximant categories (including only /w/ and /j/), inexperienced Japanese learners likely substitute the Japanese tap for English /ɪ/ in perception (Guion, Flege, Akahane-Yamada, & Pruitt, 2000) and production (Riney, Takada, & Ota, 2000). Relative to English /ɪ/, the Japanese tap sound features higher F3 (2400-3000 Hz) and F2 (1700-2100 Hz) values, and shorter transition duration (5-20 ms) (Hattori & Iverson, 2009; Lotto, Sato, & Diehl, 2004). Thus, the acquisition of English /ɪ/ pronunciation requires Japanese learners to make simultaneous constrictions in the labial, alveolar and pharyngeal parts of vocal tract (for lower F3, < 1900Hz) while retracting tongue body (for lower F2, < 1600Hz) and prolonging the phonemic length (for longer transition duration, > 50ms).

Importantly, examining this specific instance of L2 learning allowed us to probe how adult L2 learners (with different aptitude profiles) can acquire various dimensions of pronunciation abilities which are *differentially* exploited in the first language (L1) phonetic

system (i.e., existing, partially-used vs. novel cues).² According to our precursor studies (Saito & Brajot, 2013; Saito & Munro, 2014), for example, adult Japanese learners tend to show different levels of difficulty in their English /ɪ/ development in the following order: Tongue retraction/lower F2 → duration/longer F1 transition → labial, alveolar and pharyngeal constrictions/lower F3.

In the initial stages of L2 speech learning (e.g., first six months of immersion in naturalistic settings), Japanese learners tend to quickly acquire the interlanguage strategy (i.e., tongue retraction) of producing English /ɪ/ with lower F2 (< 1600 Hz) (Saito & Munro, 2014). This is arguably because this distinction (lower vs. higher F2) is used in the Japanese approximant categories (differentiating /w/ and /j/). Although lower F2 (an index of tongue retraction) is not a primary acoustic correlate of listeners' English /ɪ/ perception (for details, see below), this is the “interlanguage” strategy that Japanese learners likely adopt as a first step towards acquiring targetlike English /ɪ/ production.

After Japanese accumulate more L2 experience (a few years of immersion), they may obtain another interlanguage strategy—i.e., the ability to produce English /ɪ/ with sufficiently long length (F1 transition > 50ms) (Saito & Brajot, 2013). This is possible when Japanese learners notice, activate and integrate the length distinction which is used in the Japanese vowel categories (short vs. long vowels), but not in the Japanese approximant categories.

Finally, experienced Japanese learners demonstrate some evidence of F3 acquisition (< 1900 Hz). This typically requires a great deal of intensive L2 use (e.g., 10+ years of length of residence in an L2 speaking environment: Flege et al., 1995) or/and focused training (e.g., 30+ hours of high variability phonetic training: Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997). Producing English /ɪ/ with lower F3 (providing the most important acoustic information for English /ɪ/ perception) is considered the most difficult dimension, as the F3 acoustic representation and the relevant articulatory configuration (creating the labial, alveolar and pharyngeal constrictions) is not actively used in the Japanese system (Iverson et al., 2003).

The suggested hierarchy of Japanese learners' English /ɪ/ development also corresponds to the kinds of acoustic information native English listeners use while assessing the intelligibility and accuracy of Japanese learners' English /ɪ/ production (e.g., Iverson et al., 2003; Saito & van Poeteren, 2018; Underbakke, Polka, Gottfried, & Strange, 1988). When Japanese learners attempt to produce English /ɪ/ by using their initial interlanguage

² This study substantially drew on Flege's theoretical framework of L2 speech perception and production (Flege, 2003). Under this view, L2 speech learning is perception-based (i.e., what we can hear in an L2 will guide us to produce new sounds), and the level of learning/perceptual difficulty is determined by the extent to which “acoustic” cues are exploited in the L1 system (for more theoretical accounts of this specific view on L1-to-L2 influence, see also McAllister, Flege & Piske, 2002). Thus, using Flege's theoretical model well suits the suggested developmental stages of Japanese learners' English /ɪ/ acquisition as per the three “acoustic” categories (F2 reduction → prolonging phonemic length → F3 reduction), and thus justifies my approach towards this topic by way of an acoustic analysis of participants' L2 speech production. Note that other scholars have proposed a different idea that L2 speech learning is articulation-based (i.e., the way how we use articulators will guide us to hear new sounds); according to this view, learners are assumed to develop L2 perception and production skills simultaneously with increasing awareness of speech gestures (e.g., Best & Tyler, 2007 for Perceptual Assimilation Model-L2). Drawing on this approach would require more fine-grained analyses of L2 learners' articulation per se (e.g., ultrasound imaging). For more comprehensive overviews on various L2 speech learning theories, see Flege, 2003; Hansen Edwards & Zampini, 2008; Saito & van Poeteren, 2018.

strategy (retracting tongue body), native English listeners may have much difficulty perceiving it as English /ɪ/ because they do not use lower F2 as a perceptual cue (Iverson et al., 2003). As their experience and proficiency increases, Japanese learners likely make efforts to produce English /ɪ/ not only with more tongue retraction, but also with longer phonetic length (> 50ms). This interlanguage strategy helps native English listeners form a specific categorical perception of the sound (English /ɪ/ but not English /w/, /d/ nor /l/) (Underbakke et al., 1988). When it comes to native English listeners' *accuracy* evaluation *within* the category (the extent to which Japanese learners' English /ɪ/ pronunciation forms approximate native norms), they tend to use F3 information as a primary acoustic cue (Saito & van Poeteren, 2018).

To help native listeners perceive English /ɪ/ in an efficient and effective manner, it is reasonable to assume that Japanese learners acquire the three phonetic features at different points in time (lower F2 → longer F1 transition → lower F3) vis-à-vis increased experience and proficiency.

Current Study

The current study took an exploratory approach towards examining the aptitude-acquisition link for a specific L2 sound (i.e., English /ɪ/) among a total of 50 sophomore college Japanese students with seven years of EFL instruction and no experience overseas. In conjunction with the literature review presented above, three cognitive abilities (phonemic coding, associative memory, sequence recognition) are hypothesized to reflect three different stages of acquisition (input processing, automatization, lexicalization); and English /r/ acquisition comprises three stages of L2 speech learning according to different levels of difficulty (determined by the acoustic difference between the L1 and L2 systems) (lower F2, longer F1 duration, lower F3). Thus, the following three predictions were formulated to elucidate how different types of aptitude could be associated with different stages of English /ɪ/ development in EFL classrooms (phonemic coding for lower F2, associative memory for longer F1 duration, sequence recognition for lower F3).

1. Phonemic coding, which is essential at the initial stage of L2 learning (noticing and understanding), will relate to the extent to which Japanese learners can master the relatively easy aspect of English /ɪ/ acquisition (F2 reduction).
2. Since high-level associative memory allows L2 learners with abundant cognitive resources to process not only salient but also non-salient features in input, it was predicted that this ability would demonstrate a strong association with Japanese learners' acquisition of the medium-difficult dimension (longer duration).
3. Incidental learning aptitude, such as sequence cognition, was predicted to play a key role in the acquisition of the most difficult dimension (lower F3). This is arguably because L2 learners with this ability can maximize their L2 experience by processing the limited amount of input they receive in EFL settings intentionally as well as incidentally.

Participants

Japanese EFL Students. The 50 sophomore EFL students (range = 18-19 years; 23 males, 27 females) were recruited from various social sciences and humanities programs (business, marketing, economics, psychology) at a large university located in downtown Tokyo. The participants were carefully selected to ensure that they had relatively homogeneous EFL backgrounds according to the following criteria. First, they were native speakers of Japanese (both of their parents must have been L1 Japanese speakers). Second, they started learning English from Grade 7, that is, they had seven years of EFL education prior to entering the university.³ Third, they had never been abroad at the time of the project. In this regard, we assumed that individual differences in their L2 pronunciation performance could be accounted for by their supposedly different aptitude profiles but not by their relatively similar experience.⁴

Preliminary analyses of the same participants' general oral proficiency (elicited from a picture cartoon task) and aptitude scores (based on LLAMA-B, D, E, and F) were reported in Saito (2017). In the current study, the same participants engaged in three different tasks (word reading, sentence reading, timed picture description), where their abilities to pronounce English /ɪ/ in particular were tested at both controlled and spontaneous speech levels. Subsequently, the three different aspects of their English /ɪ/ pronunciation (F3, F2, duration) were acoustically analyzed and linked to their relevant aptitude profiles of phonemic coding (LLAMA-E), associative memory (LLAMA-B) and sequence recognition (LLAMA-D). Within the first four weeks of the first semester in Year 2, the participants came for individual data collection sessions, which comprised of an interview (to survey their bio and previous EFL experience), speech recording (to examine their English /ɪ/ pronunciation) and aptitude test (to survey their explicit and incidental pronunciation learning abilities). The entire session lasted for approximately one hour.

Native Baselines. In addition, to provide baseline data for English /ɪ/ production, a total of 10 native speakers of English were also recruited (5 males, 5 females) from a university in Vancouver, Canada ($M_{age} = 25.2$ years). They were considered to be “monolinguals,” given that they reported little familiarity/use of other foreign languages on a daily basis.

Measuring Aptitude Profile

To measure the participants' aptitude, three subcomponents of the LLAMA aptitude test—LLAMA-D (sequence recognition), LLAMA-B (associative memory) and LLAMA-E (phonemic coding)—were adopted in the current study for the following reasons. According to Skhan's (2016) taxonomy on L2 aptitude tests, the LLAMA is considered as domain-

³ After their entrance to the university, and at the time of the project, all the participants were required to take one foreign language class of their choice (Chinese, French, German, Spanish) which lasted for three hours per week.

⁴ Despite the participants' identical length of EFL learning (i.e., seven years), they could have been different in other respects, such as the quality of experience (how intensively they practiced English within seven years of EFL education), motivation, beliefs, emotion and personality. This methodological point/limitation will be revisited in the Discussion section.

specific. The test consists of linguistic materials and tasks which simulate various kinds of L2 learning experiences when participants aim to acquire a new, unfamiliar language. The methodological decision here distinguished the current study from the previous aptitude and speech studies using non-verbal materials (e.g., Darcy et al., 2015).

Additionally, the three tasks in the LLAMA (sequence recognition, associative memory, phonemic coding) correspond to different cognitive abilities particularly relevant to L2 pronunciation learning—i.e., phonemic coding for input processing; associative memory for automatization; and sequence recognition for lexicalization (Skehan, 2016).

Finally, the LLAMA enables us to highlight two different modes of participants' L2 learning processes—explicit vs. incidental. From a methodological point of view, the three subtests—LLAMA-D, B, E—differ in terms of the absence/presence of the practice phase, where participants explicitly study the test materials. In this regard, the LLAMA-D was presumed to tap into incidental sound learning ability, and the LLAMA-B and LLAMA-E into intentional sound learning ability. To avoid activating the participants' awareness, they engaged in the subtests in the following order: LLAMA-D → LLAMA-B → LLAMA-E.

- **LLAMA-D.** First, the participants listened to 10 sound strings as a part of a sound check (instructed to do so in order to avoid invoking any “intention” to memorize or analyze new sounds). They were then immediately asked to proceed to the testing phase,⁵ where they listened to 30 new sound strings, and answered whether they had heard each item during the first listening. Their scores were recorded out of 75 points.
- **LLAMA-B.** Different from the LLAMA-D, in this subtest the participants were explicitly asked to remember a combination of five-to-six letters for 20 imaginary objects (i.e., associative memory) within two minutes so as to prepare for the testing phase afterwards (recollection of learned items). Their scores were recorded out of 100 points.
- **LLAMA-E.** In this subtest, participants were given two minutes to identify and remember the relationship between 24 unfamiliar sounds and symbols (one syllable per symbol). Subsequently, their understanding/memory of sound-symbol correspondence was tested while listening to 20 combinations of two-sound strings. Their scores for the LLAMA-E were recorded out of 100 points.

Measuring English /ɹ/ Pronunciation

The same English /ɹ/ analysis procedure in the precursor research (for the details, see Saito & Brajot, 2013) was adopted. First, the participants engaged in three different tasks—Timed Picture Description (TPD), Sentence Reading (SR) and Word Reading (WR)—which require different processing abilities (spontaneous vs. controlled). Afterwards, the participants' English /ɹ/ tokens (four tokens [*read, rain, rock, road*] from TPD; eight tokens

⁵ According to the casual interview after the task, the participants reported that they fully focused on checking their hearing abilities (without any intention to remember the sound strings), as they were told to do so.

[*read, rain, red, race, run, Ryan, road, wrong*] from SR; eight tokens [*read, room, root, red, race, rough, ram, right*] from WR) were acoustically analyzed according to F2, F1 transition and F3, each of which served as an index for the degree and rate of tongue retraction and labial/alveolar/pharyngeal constrictions.

Stimuli. All 20 target words included /ɹ/ in the word-initial position ($n = 8$ for WR, $n = 8$ for SR, $n = 4$ for TPD) and were Consonant-Vowel-Consonant (CVC) singletons except “*Ryan*” in SR (CVVC). Given that Japanese learners tend to have more difficulty in producing word-initial /ɹ/ preceding front vowels (e.g., *read, rain*) than central and back vowels (*rock, road*) (Flege et al., 1995), the following vowel condition was carefully controlled in each task (50% for singletons with front vowels and 50% for singletons with central and back vowels). The test tokens are summarized in Table 1.

Table 1

N = 20 English /ɹ/ Tokens in the Controlled and Spontaneous Production Tests in Relation to Following Vowel Conditions

A. Timed Picture Description	
<i>Following vowels</i>	
[front]	read, rain
[central/back]	road, rock
B. Sentence Reading	
<i>Following vowels</i>	
[front]	read, rain, red, race
[central/back]	run, Ryan, road, wrong
C. Word Reading	
<i>Following vowels</i>	
[front]	read, red, race, ram
[central/back]	rough, right, root, room

Procedure. First, the participants engaged in TPD. To elicit participants’ spontaneous use of English /ɹ/, they described seven pictures with five seconds of planning time for each photo. While the first three photos served as distractors (to reduce task familiarity effects), the remaining four photos were used for the main analysis. Three key words were given for each photo that the participants had to use while completing the task; one of the key words was a target word. For example, one picture portrayed a table in the drive way under the rain (key words: *table, driveway, rain*). The purpose of this picture was to elicit learners’ spontaneous production of /ɹ/ in *rain*.

Second, participants read five target sentences together with three distracter sentences. A total of eight target words including word-initial /ɹ/ were embedded in the target sentences. In this sentence reading task, the participants’ abilities to pronounce English /ɹ/ under more controlled conditions (than TPD) were tested. The target sentences were as follows:

- *He will read my paper by the time I arrive there.*
- *She left her red bicycle on the side of the road.*

- *The race was cancelled because of the rain.*
- *I can correct all wrong sentences tonight.*
- *Ryan does not like to run in the snow.*

Finally, participants read a list of 25 individual words containing eight target words (*read, room, root, red, race, rough, ram, and right*) and 17 distracters (see Appendix). Different from TPD (where participants' priority lied in conveying intended message), they were asked to pronounce each word as accurately as possible (with their focus on accurate pronunciation forms).

Individual recording sessions took place in a sound proof booth at the Japanese university (for the Japanese EFL participants) and at a university-level school in Vancouver (for the native baselines). To avoid any confusion/misunderstanding in the oral test procedure for the Japanese participants, all instructions were delivered in Japanese by a research assistant (a native speaker of Japanese). All speech tokens were recorded by way of a Marantz PMD 660 and Roland-05, set at 44.1 kHz sampling rate and 16-bit quantization.

Acoustic analysis. In conjunction with the procedure for the acoustic analysis of natural English /ɪ/ tokens used in Flege et al. (1995) and Hattori & Iverson (2009), two spectral (F3, F2) and temporal (F1 transition) aspects of English /ɪ/ production was analyzed through a linear predictive coding routine in *Praat* (Boersma & Weenik, 2017). For WR, word onset was identified by looking at both the wave forms and spectrographic representation of each token. Then, a cursor was placed on the point where both F2 and F3 were clearly observed. For TPD, a cursor was placed on the local minimum of F3 (dip) to get the F2 and F3 values. The length of English /ɪ/ was measured by measuring the beginning and end of F1 transition.

Normalization. To normalize anatomical differences among the participants (e.g., vocal tract length), their raw F2 and F3 values were adjusted following Lee, Guion, and Harada's (2006) procedure (for details, see also Saito & Brajot, 2013; Saito & Munro, 2014).

Results

Aptitude Scores

The participants' scores on the LLAMA-D, LLAMA-B and LLAMA-E are summarized in Table 2. To examine the construct validity of the LLAMA test, the independence/dependence of the participants' scores were checked via correlation analyses. According to the results of the Kolmogorov–Smirnov goodness-of-fit tests, all the aptitude scores followed the normal distribution ($p > .05$).

Table 2

Descriptive Statistics of Aptitude Test Scores by 50 Japanese Students

	<i>M</i>	<i>SD</i>	Range		95% CI	
			Min	Max	Lower	Upper
<u>Language aptitude</u>						
LLAMA-D (75 points)	41.8	13.2	10	70	38.1	45.4
LLAMA-B (100 points)	60.0	17.3	25	100	55.2	64.7
LLAMA-E (100 points)	83.4	19.4	10	100	77.8	88.7

The participants' LLAMA-D, LLAMA-B and LLAMA-E scores were submitted to Pearson correlation analyses with alpha set to $p < .017$ (Bonferroni corrected). The three aptitude subtest scores were not significantly related to each other ($r = .05$, $p = .70$ for LLAMA-D vs. LLAMA-B; $r = -.06$, $p = .67$ for LLAMA-D vs. LLAMA-E; $r = .20$, $p = .14$ for LLAMA-B vs. LLAMA-E). As conceptualized by Meara (2005), these subtest scores could thus be considered to tap into three different dimensions of the participants' cognitive abilities for language learning: (a) sequence recognition ability (LLAMA-D), associative memory ability (LLAMA-B) and phonemic coding ability (LLAMA-E).

English /ɪ/ Proficiency

According to the descriptive results (see Table 3), the Japanese students' /ɪ/ performance was substantially different from the native baselines in all the acoustic dimensions (F2, F1 transition duration, F3). The Kolmogorov–Smirnov goodness-of-fit tests confirmed the normal distribution of the participants' F2, F1 transition duration and F3 values ($p > .05$). As for the task dimensions, the results of one-way ANOVA showed that the acoustic dimensions of their pronunciation forms did not significantly differ between WR, SR and TPD in terms of F2, $F(2, 110) = 1.988$, $p = .101$, duration, $F(2, 110) = 1.638$, $p = .119$, and F3, $F(2, 110) = 2.434$, $p = .092$. The results indicated that the Japanese participants' English /ɪ/ performance was comparable whether it was elicited via the controlled or spontaneous tasks (WR, SR vs. TPD).

Table 3

Descriptive Results of English /ɪ/ Production by Japanese Students (n = 50), and Native Baselines (n = 10)

		Japanese students (n = 50)		Native Baseline (n = 10)	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
F2	WR	1627	271	1319	177
	SR	1595	245	1310	132
	TPD	1659	258	1369	184
Duration	WR	35.4	22.9	91.9	10.5
	SR	35.5	17.6	80.9	11.0
	TPD	31.3	17.5	89.2	12.7
F3	WR	2629	455	1928	113
	SR	2607	398	1931	106
	TPD	2680	358	1923	84.4

Notes. WR for word reading; SR for sentence reading; TPD for timed picture description.

To count how many Japanese participants' performance could be considered within the range of English /ɪ/, the nativelikeness analysis was adopted from many naturalistic L2 ultimate attainment studies (e.g., Abrahamson, 2012). Following the commonly-used procedure of the analysis, certain Japanese learners could be regarded as having reached "near-nativelike" proficiency if their performance fell within two SDs of the native baseline group's average scores. Accordingly, the number of such Japanese students who reached the native threshold of English /ɪ/ was counted for all the acoustic dimensions (F2, duration, F3) at WR, SR and TPD, respectively.

Table 4

No. of Japanese Students Who Fell within the Range of nativelike English /ɪ/

		<i>n</i>	<i>% (out of N = 50)</i>	<i>Threshold^a</i>
F2	WR	28	56%	< 1674 Hz
	SR	27	54%	< 1574 Hz
	TPD	21	42%	< 1637 Hz
Duration	WR	4	8%	> 70.6 ms
	SR	7	14%	> 58.8 ms
	TPD	1	2%	> 63.7 ms
F3	WR	9	18%	< 2156 Hz
	SR	5	10%	< 2143 Hz
	TPD	1	2%	< 2092 Hz

Notes. ^aAll the participants whose performance was below/above the threshold could be considered "nativelike" for each category (F2, duration, F3). WR for word reading; SR for sentence reading; TPD for timed picture description

As summarized in Table 4, approximately half of the participants (42-56%) acquired the nativelike F2 representation for English /ɪ/. However, the ratio of such successful English /ɪ/ performance was relatively low in the acquisition of duration (2-14%) and F3 (2-18%). The results indicated that seven years of EFL education led many Japanese learners to master the easy aspect of English /ɪ/ acquisition (F2). However, much individual variability was present as to the difficult aspect of English /ɪ/ acquisition (duration, F3).

Aptitude-Proficiency Link

To provide a general picture of how each dimension of the participants' English /ɪ/ pronunciation—F2, F1 transition duration and F3—was associated with their sequence recognition (LLAMA-D), associative memory (LLAMA-B) and phonemic coding (LLAMA-E), a set of Pearson correlation analyses were performed. The size of the associations was also considered in consultation with Plonsky and Oswald's (2014) field-specific benchmarks ($r = .25$ for small, $.40$ for medium, $.60$ for large).

As shown in Table 5, the results identified significant correlations between the participants' F2 performance at WR and LLAMA-E ($p = .019$). In addition, LLAMA-B was marginally related to the participants' duration at WR ($p = .052$), and significantly correlated with their F2 and F3 performance at WR ($p = .027, .046$). The strength of the significant/marginal contrasts here ($r = |.277$ to $-.312$) were considered small-to-medium (Plonsky & Oswald, 2014).

Table 5

Correlations between the Participants English /ɪ/ Performance and Aptitude Scores

		LLAMA-D		LLAMA-B		LLAMA-E	
		<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
F2	WR	-.192	.181	-.312	.027*	-.330	.019*
	SR	-.140	.332	-.200	.163	-.189	.188
	TPD	-.189	.189	-.232	.105	-.140	.333
Duration	WR	-.078	.590	.277	.052†	.178	.216
	SR	-.186	.195	.070	.627	.001	.997
	TPD	-.177	.219	.099	.492	.110	.448
F3	WR	.051	.725	-.283	.046*	-.172	.233
	SR	.156	.280	-.143	.320	.007	.960
	TPD	.053	.713	-.102	.480	-.039	.787

Note. *indicates statistical significance at $p < .05$ †indicates marginal significance at $p < .10$

To further probe the predictive power of the participants' different aptitude scores (sequence recognition, associative memory, phonemic coding) for three different aspects of their English /ɪ/ performance (F2, F1 duration, F3), a set of stepwise multiple regression analyses were performed with each acoustic dimension of English /ɪ/ as dependent variables relative to three different aptitude scores as independent variables. As summarized in Table 6, LLAMA-E scores explained 10.9% of the variance in the F2 performance; and LLAMA-B accounted for and 8.0% of the variance in the F3 performance. According to Plonsky and

Oswald’s (2014) benchmark ($R^2 = .06$ for small, $.16$ for medium, $.36$ for large), the amount of explained variance in these models could be considered relatively “small-to-medium” ($.06 < R^2 < .16$).

Table 6

Results of Multiple Regression Analyses Using the Aptitude Scores as Predictors of English /ɪ/ Pronunciation Performance

Acoustic dimensions	Task conditions	Predictor variables	<i>R</i>	<i>Adjusted R²</i>	<i>F</i>	<i>p</i>
F2	WR	LLAMA-E	.330	.109	5.876	.019
	SR			<i>n.s.</i>		
	TPD			<i>n.s.</i>		
Duration	WR			<i>n.s.</i>		
	SR			<i>n.s.</i>		
	TPD			<i>n.s.</i>		
F3	WR	LLAMA-B	.283	.080	4.192	.046
	SR			<i>n.s.</i>		
	TPD			<i>n.s.</i>		

Notes. WR for word reading; SR for sentence reading; TPD for timed picture description; LLAMA-B for associative/rote memory; LLAMA-E for phonemic coding

Discussion

In the context of 50 second-year college Japanese students with relatively homogeneous EFL backgrounds (i.e., seven years of EFL education and no experience abroad), the current study examined the extent to which individual variability in English /ɪ/ pronunciation attainment could be explained by three different components of aptitude—sequence recognition (LLAMA-D), associative memory (LLAMA-B) and phonemic coding (LLAMA-E). In keeping with Skehan’s (2016) proposal, these aptitude factors were hypothesized to facilitate different stages of L2 acquisition (phonemic coding for initial, associative memory for mid, sequence recognition for advanced; Skehan, 2016). Furthermore, the participants’ English /ɪ/ performance was scrutinized according to different acoustic/task dimensions which were assumed to reflect different amounts of learning difficulty (lower F2 for tongue retraction < longer duration for prolonged phonemic length < lower F3 for labial, alveolar and pharyngeal constrictions) and processing abilities (Word Reading < Sentence Reading < Timed Picture Description).

First and foremost, the results of the acoustic analyses showed a great deal of individual variability in the Japanese participants’ English /ɪ/ performance after seven years of EFL education. Whereas many of them (40-60%) successfully attained nativelike proficiency for the relatively easy aspect of English /ɪ/ acquisition ($F2 < 1700$ Hz), very few (< 20%) did so for the relatively difficult aspect of English /ɪ/ acquisition (duration > 50ms, $F3 < 2100$ Hz). This was consistent with major L2 speech learning theories (e.g. Flege, 2003 for Speech Learning Model) which equally state that this specific L1-L2 context—the acquisition of English /ɪ/ by adult Japanese learners—could be considered as one of the most

difficult instances of sound learning. Indeed, the available research evidence has thus far suggested that L2 pronunciation development is a slow, gradual phenomenon (requiring more than 10 years of immersion for acquiring all aspects of English /ɪ/ pronunciation: see Ingvalson, McClelland, & Holt, 2011). Taken together with what was found in the current study, it is logical to speculate that seven years of EFL education alone may not be sufficient for Japanese students to lead to robust pronunciation development for all the acoustic dimensions of English, especially without any immersion through study-abroad (Riney & Flege, 1998).

Given that the amount of successful L2 pronunciation learning may widely vary regardless of learners' EFL experience factors, the question has now become: To what degree can such individual variability in classroom L2 pronunciation learning be explained by factors related to aptitude? As predicted earlier, the results identified explicit learning aptitude—phonemic coding, associative memory—as a significant affecting factor for the participants' English /ɪ/ acquisition. Whereas those with higher phonemic coding abilities appeared to demonstrate better performance in the relatively easy dimension of English /ɪ/ pronunciation (lower F2), those with greater associative memory tended to demonstrate more advanced performance, especially in the relatively difficult dimension of English /ɪ/ pronunciation (longer transition duration, lower F3). In line with the field-specific benchmark, suggested by Plonsky and Oswald (2014), the magnitude of the associations could be considered “small-to-medium.”

Extending Skehan's (2016) acquisition-aptitude framework (which has evolved in L2 morphosyntax studies), the findings here indicated a multifaceted relationship between different constructs of aptitude and different stages of L2 pronunciation development.

Phonemic Coding

Some scholars have shown that phonemic coding could be linked to global constructs of L2 pronunciation learning (e.g., Granena & Long, 2013 for foreign accentedness). To our knowledge, the current study was the very first attempt to examine the role of phonemic coding ability in the specific instance of L2 segmental acquisition with different levels of learning difficulty—Japanese learners' F2, F3 and duration of English /ɪ/ pronunciation.

Phonemic coding ability, which requires both analysis (identifying the underlying patterns between sound strings to alphabets with diacritics) and memory (remembering sound-symbol correspondence) aspects of cognition may allow L2 learners to briefly hold onto incoming input and make it available for quick and immediate phonetic analysis. When exposed to a new target sound (English /ɪ/), such talented L2 learners may quickly notice, understand and adjust to the already-existing acoustic (lower F2) and articulatory (tongue retraction) parameters in the L1 Japanese phonetic system (where F2 is used to differentiate between /w/ and /j/). However, phonemic coding did not show significant correlations with the acquisition of the relatively difficult aspects of English /ɪ/ learning—phonemic lengthening and F3 reduction.

Associative Memory

In the previous literature, it has been shown that Japanese learners tend to work on various aspects of English /ɪ/ production over an extensive period of time in naturalistic

settings (length of residence > 10 years). To achieve this demanding goal under classroom conditions, where input is relatively limited, Japanese learners may need more robust memory abilities, such as associative memory, to promote the acquisition of the phonetic cues absent in the L1 Japanese phonetic system, and to enhance the intelligibility and accuracy of English /ɪ/ pronunciation. Associative memory is hypothesized to help L2 learners better perceive and produce the target sound. Such talented learners can store larger amounts of visual, written and sound information, which will, as a result, free up more cognitive resources available for more detailed, deeper analyses of the partially existing cues (longer phonemic length), as well as new cues (lower F3 for labial, alveolar and pharyngeal constrictions).

Our findings on the relationship between associative memory and the later stage of L2 speech learning are in line with the previous aptitude literature, which has shown that L2 learners with greater associative memory likely attain advanced L2 proficiency in terms of grammaticality judgements (Linck et al., 2013) and reading and listening skills (Schneiderman & Desmarai, 1988). As for L2 speech learning, Silbert et al. (2015) examined and confirmed the strong relationship between associative memory and perception of non-native contrasts.

For theoretical relevance, the findings can also be interpreted with reference to the influencing working memory model in cognitive psychology (e.g., Baddeley, 2003). In fact, there has been a recent paradigm shift in the field of SLA towards integrating working memory model into a component of foreign language aptitude (e.g., Linck, Osthus, Koeth, & Bunting., 2014). According to Li's (2016) meta-analysis, the construct of the "short-term store" component of working memory significantly overlapped with the construct of associative memory (whereas the "executive control" component of working memory was closely tied to the construct of phonemic coding). This is arguably because of the fact that tasks measuring both associative memory (paired associates) and phonological short-term memory (non-word repetition, non-word/digit span) are similar in nature—i.e., storing and rehearsing new verbal/non-linguistic information in phonological code but without entailing much processing.

If we take the stance that associative memory is possibly linked to phonological short-term memory, the findings of the study (i.e., the role of associative memory in the relatively difficult aspects of English /ɪ/ acquisition) would provide another empirical evidence to the strong memory effects on high-level L2 acquisition, reported in a wide range of SLA studies to date. For example, L2 learners with greater phonological short-term memory likely attain more nativelike lexicogrammar performance as a result of study abroad (O'Brien, Segalowitz, Collentine, & Freed, 2006) and long-term immersion (Foster, Bolibaugh, & Kotula, 2014). By conducting an intervention study with a pretest-and-posttest design, Révész (2012) demonstrated that the impact of instruction (recasts) on L2 oral proficiency development (the accurate use of past/present progressive) was significantly correlated with participants' phonological short-term memory (but not with their executive control). Echoing the SLA/working memory literature (see Kormos, 2013 for a detailed review), the current study supports a tentative conclusion that solid phonological memory could be instrumental to advanced-level L2 speech attainment, because it is hypothesized to equip learners with

strong storage functions which they need so as to best utilize, process and analyze every L2 input that they have explicitly engaged in.

Roles of Incidental Learning Aptitude

Finally, it is important to remember that the participants' incidental aptitude—sequence recognition—was not associated with their English /ɪ/ performance, and that the link between explicit aptitude—phonemic coding, associative memory—and all the significant aptitude-proficiency associations were limited to the controlled task condition (word reading). The lack of any significant role for incidental aptitude (sequence recognition) hints that incidental learning could be minimally beneficial in classroom settings, where the nature of L2 learning is predominantly explicit (e.g., grammar translation, form-focused exercise activities). Under such L2 learning environments, as shown earlier, L2 learners seemingly benefit from those components of aptitude which impact explicit and intentional learning in an effective and efficient manner. Similarly, the explicit aptitude effects found in the controlled task (WR) rather than the spontaneous task (TPD) also indicate that such aptitude can predict L2 learners' performance when they can fully focus on phonetic forms without paying much attention to the content of message (controlled processing)—a typical characteristic of many form-oriented EFL classrooms.

Future Directions

Given that the current study took a first step towards conceptualizing an aptitude framework for L2 pronunciation development in classroom settings, the interpretations of the dataset presented here need to be considered as tentative at best. Given the theoretical and practical value of the topic, several issues worthy of future investigation need to be addressed. First, although the study found that the two explicit aptitude factors explained approximately 10% of variance in the participants' L2 pronunciation attainment, it would be intriguing to replicate the findings by adopting a more comprehensive set of aptitude measures which are designed to tap into explicit and implicit learning abilities (with and without awareness), and domain-specific and -general abilities (associated with language learning and applicable to any general skill acquisition) in both the receptive and productive modes (cf. Saito, Sun, & Tierney, in press for explicit and implicit phonemic awareness)

Furthermore, the current study assumed that the participants had relatively homogeneous EFL backgrounds at the time of the project (7 years of EFL experience without any experience abroad). Although the participants' aptitude scores were moderately associated with their English /ɪ/ pronunciation, there is some possibility that the findings here could have been confounded with their potentially different EFL backgrounds (e.g., those who attained better English /ɪ/ pronunciation may have had not only higher aptitude scores but also spent more time in practicing the L2). On the one hand, it has been extensively shown that adult L2 learners can demonstrate some improvement (e.g., their pronunciation forms being more intelligible) after receiving a great amount of L2 input under EFL conditions (5-6 years of classroom experience) (Simon & D'Hulster, 2012).

On the other hand, few studies have thus far examined precisely how many hours and what kind(s) of L2 instruction are needed to enhance the rate and ultimate attainment of L2 pronunciation. While tracking the longitudinal development of L2 oral proficiency by

Japanese EFL students, Saito and Hanzawa (2017) showed that the effects of classroom experience on Japanese EFL students' change could be limited to fluency (rather than segmental) aspects of L2 speech. In terms of Japanese learners' English /ɪ/ acquisition, Riney and Flege's (1998) longitudinal study similarly demonstrated that the mere exposure to L2 instruction alone may not result in tangible improvement, suggesting that some learner-internal factors beyond experience (e.g., aptitude) may predict the incidence of successful L2 speech learning in foreign language classrooms. To this end, future research is strongly called for with a view of expounding the complex relationship between EFL experience, aptitude and individual differences vis-à-vis the context of L2 segmental development and attainment (cf. Nagle, 2017).

Last, the current study, which used a cross-sectional dataset, does not allow us to discuss any causal relationship between aptitude and proficiency. As a remedy, more longitudinal work is needed to further examine the extent to which L2 learners with special aptitude profiles (associative memory, phonemic coding, phonological short-term memory) can actually enhance various dimensions of L2 speech, when they engage in various kind of practice activities over an extensive period of time (e.g., O'Brien et al., 2006).

References

- Abrahamsson, N. (2012). Age of onset and nativelike L2 ultimate attainment of morphosyntactic and phonetic intuition. *Studies in Second Language Acquisition*, 34, 187–214.
- Alderson, J. C., Clapham, C., & Steel, D. (1997). Metalinguistic knowledge, language aptitude and language proficiency. *Language Teaching Research*, 1, 93–121.
- Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, 4, 829-839.
- Boersma, D., & Weenink, P. (2017). *Praat: Doing phonetics by computer version 6.0.29*. Retrieved from <http://www.praat.org>
- Bradlow, A., Pisoni, D., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/. *Journal of the Acoustical Society of America*, 101, 2299–2310.
- Carroll, J.B. (1962). The prediction of success in intensive foreign language training. In R. Glaser (Ed.), *Training, research, and education* (pp. 131-151). New York: Wiley.
- Carroll, J. B., & Sapon, S. M. (1959). *Modern language aptitude test*.
- Darcy, I., Park, H., & Yang, C. L. (2015). Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing. *Learning and Individual Differences*, 40, 63-72.
- Darcy, I., Mora, J., & Daidone, D. (2016). The role of inhibitory control in second language phonological processing. *Language Learning*. Advanced online publication. doi: 10.1111/lang.12161
- Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics*, 27, 1–24.
- Espy-Wilson, C.Y. (1992). Acoustic measures for linguistic features distinguishing the semivowels /w j r l/ in American English. *Journal of the Acoustical Society of America*, 92, 736–757.
- Espy-Wilson, C.Y., Boyce, S.E., Jackson, M., Narayanan, S. & Alwan, A. (2000). Acoustic modeling of American English /ɹ/. *Journal of the Acoustical Society of America*, 108, 343–356.
- Flege, J. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer & N. Schiller (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 319–355). Berlin: Mouton de Gruyter.
- Flege, J., Takagi, N., & Mann, V. (1995). Japanese adults learn to produce English /ɹ/ and /l/ accurately. *Language and Speech*, 38, 25–55.
- Foster, P., Bolibaugh, C., & Kotula, A. (2014). Knowledge of native selections in a L2: The Influence of exposure, memory, age of onset, and motivation in foreign language and immersion settings. *Studies in Second Language Acquisition*, 36, 101-132.
- Granena, G. (2013b). Individual differences in sequence learning ability and second language acquisition in early childhood and adulthood. *Language Learning*, 63, 665–703.
- Guion, S., Flege, J., Ahahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults'

- perception of English consonants. *Journal of the Acoustical Society of America*, 107, 2711–2725.
- Hansen Edwards, J. G., & Zampini, M. L. (Eds.). (2008). *Phonology and second language acquisition*. Amsterdam: John Benjamins.
- Hattori, K., & Iverson, P. (2009). English /ɪ/-/i/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *Journal of the Acoustical Society of America*, 125, 469–479.
- Ingvalson, E. M., McClelland, J. L., & Holt, L. L. (2011). Predicting native English-like performance by native Japanese speakers. *Journal of phonetics*, 39, 571-584.
- Iverson, P., Kuhl, P., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87, 847–857.
- Kormos, J. (2013). New conceptualizations of language aptitude in second language attainment. In G. Granena & M. Long (Eds.), *Sensitive periods, language aptitude and ultimate attainment* (pp. 131-152). Amsterdam: John Benjamins.
- Li, M., & DeKeyser, R. (2017). Perception practice, production practice, and musical ability in L2 Mandarin tone-word learning. *Studies in Second Language Learning*. Advanced online publication. doi: 10.1017/S0272263116000358
- 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.
- Li, S. (2015a). The Associations between language aptitude and second language grammar acquisition: A meta-analytic review of five decades of research. *Applied Linguistics*, 36, 385-408.
- Li, S. (2015b). The differential roles of two aptitude components in mediating the effects of two types of feedback on the acquisition of an opaque linguistic structure. In C. Sanz and B. Lado (Eds.) *Individual differences, L2 development, and language program administration: From theory to application* (pp. 32-52). Boston, MA: Cengage Learning.
- Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition*. doi: 10.1017/S027226311500042X
- Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., & Doughty, C. J. (2013). Hi-LAB: A New Measure of Aptitude for High-Level Language Proficiency. *Language Learning*, 63, 530-566.
- Linck, J., Osthus, P., Koeth, J., & Bunting, M. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, 21, 861-883.
- Lee, B., Guion, S.G., & Harada, T. (2006). Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals. *Studies in Second Language Acquisition* 28, 487–513.
- Lotto, A. J., Sato, M., & Diehl, R. L. (2004). Mapping the task for the second language learner: The case of Japanese acquisition of /ɪ/ and /i/. In J. Slifka, S. Manuel, & M. Matthies (Eds.), *From sound to sense: 50+ years of discoveries in speech*

communication (pp. C381–C386). Cambridge, MA: Research Laboratory of Electronics at MIT.

- Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics*, 35, 463–482.
- Nagle, C. (2017). A longitudinal study of voice onset time development in L2 Spanish stops. *Applied Linguistics*. DOI: 10.1093/applin/amx011
- O'Brien, I., Segalowitz, N., Collentine, J., & Freed, B. (2006). Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics*, 27, 377-402.
- Révész, A. (2012). Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning*, 62, 93-132.
- Plonsky, L., & Oswald, F. L. (2014). How big is 'big'? Interpreting effect sizes in L2 research. *Language Learning*, 64, 878-912.
- Riney, T., & Flege, J. (1998). Change over time in global foreign accent and liquid identifiability and accuracy. *Studies in Second Language Studies*, 20, 213–243.
- Riney, T., Takada, M., & Ota, M. (2000). Segmentals and global foreign accent: The Japanese flap in EFL. *TESOL Quarterly*, 34, 711–737.
- Saito, K. (2017). Effects of sound, vocabulary and grammar learning aptitude on adult second language speech attainment in foreign language classrooms. *Language Learning*, 67, 665-693.
- Saito, K., & Brajot, F. (2013). Scrutinizing the role of length of residence and age of acquisition in the interlanguage pronunciation development of English /ɪ/ by late Japanese bilinguals. *Bilingualism: Language and Cognition*, 16, 847-863.
- Saito, K., & Hanzawa, K. (2017). The role of input in second language oral ability development in foreign language classrooms: A longitudinal study. *Language Teaching Research*. DOI: 10.1177/1362168816679030
- Saito, K., & Munro, M. (2014). The early phase of /r/ production development in adult Japanese learners of English. *Language and Speech*, 57, 451-469.
- Saito, K., Sun, H., & Tierney, A. (in press). Explicit and implicit aptitude effects on second language speech learning: Scrutinizing segmental, prosodic and temporal sensitivity and performance via behavioral and neurophysiological measures *Bilingualism: Language and Cognition*.
- Saito, K., Suzukida, Y., & Sun, H. (2018). Aptitude, experience and second language pronunciation proficiency development in classroom settings: A longitudinal study. *Studies in Second Language Acquisition*. DOI: 10.1017/S0272263117000432
- Saito, K. & van Poeteren, K. (2018). The perception-production link revisited: The case of Japanese learners' English /r/ performance. *International Journal of Applied Linguistics*, 28, 3-17.
- Schneiderman, E. I. & Desmarais, C. (1988). The talented language learner: Some preliminary findings. *Second Language Research*, 4, 91–109.
- Silbert, N. H., Smith, B. K., Jackson, S. R., Campbell, S. G., Hughes, M. M., & Tare, M. (2015). Non-native phonemic discrimination, phonological short term memory, and word learning. *Journal of Phonetics*, 50, 99-119.

- Simon, E., & D'Hulster, T. (2012). The effect of experience on the acquisition of a non-native vowel contrast. *Language Sciences*, 34, 269–283.
- Skehan, P. (2015). Foreign language aptitude and its relationship with grammar: A critical overview. *Applied Linguistics*, 36, 367–384.
- Skehan, P. (2016). Foreign language aptitude, acquisitional sequences, and psycholinguistic processes. In G. Granena, D. Jackson & Y. Yilmaz (Eds.), *Cognitive individual differences in L2 processing and acquisition* (pp. 15–38). Amsterdam: John Benjamins
- Trofimovich, P., Kennedy, S., & Foote, J. A. (2015). Variables affecting L2 pronunciation development. In M. Reed & J. Levis (Eds.), *The handbook of English pronunciation* (pp. 353-373). Malden, MA: Wiley.
- Underbakke, M., Polka, L., Gottfried, T. L., & Strange, W. (1988). Trading relations in the perception of /ɹ/ and /l/ by Japanese learners of English. *Journal of the Acoustical Society of America*, 84, 90–100.
- Yalçın, Ş., & Spada, N. (2016). Language aptitude and grammatical difficulty. *Studies in Second Language Acquisition*, 38, 239-263.
- Yilmaz, Y. & Koşlu, Y. (2016). The interaction between phonetic coding ability and feedback exposure condition. In G. Granena, D. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in L2 processing and acquisition* (pp. 303-326). Amsterdam: John Benjamins.
- Yilmaz, Y., & Grañena, G. (2016). The role of cognitive aptitudes for explicit language learning in the relative effects of explicit and implicit feedback. *Bilingualism: Language and Cognition*, 19, 147-161.

Appendix: Distracter Items in Word Reading

man, book, desk, tall, bus, music, Tom, ship, chair, map, mom, sip, subway, yellow, think,
feet, cap