



The Role of Input in Second Language Oral Ability Development in Foreign Language Classrooms: A Longitudinal Study

Kazuya Saito (Birkbeck, University of London)

Keiko Hanzawa (Waseda University)

Abstract

The current project longitudinally investigated the extent to which first-year Japanese university students developed their second language (L2) oral ability in relation to increased input in foreign language classrooms. Their spontaneous speech was elicited at the beginning, middle and end of one academic year, and then judged by linguistically trained coders for pronunciation, fluency, vocabulary and grammar qualities. According to the statistical analyses, the total amount of input (operationalized as number of English classes taken and L2 use outside of classrooms) was significantly related to the participants' quick and immediate development of fluency and lexicogrammar during the first semester. Their pronunciation development was mixed, either subject to continuous change over two academic semesters (for prosody) or limited within the timeframe of the study (for segmentals). Similar to naturalistic L2 speech learning, the findings support the multifaceted role of input in different areas of oral proficiency development in foreign language classrooms.

Key words: Second language speech, Foreign language education, Pronunciation, Fluency, Vocabulary, Grammar

Title:

The Role of Input in Second Language Oral Ability Development in Foreign Language
Classrooms: A Longitudinal Study

Running Head:

FOREIGN LANGUAGE SPEECH LEARNING

Acknowledgement

This study was funded by the Grant-in-Aid for Scientific Research in Japan (No. 26770202, No. 244722). We are grateful to anonymous *LTR* reviewers and the journal editor, Frank Boers, for their helpful input and feedback on the content of this manuscript, and to George Smith and Ze Shan Yao who helped data analyses. Finally, we thank Pavel Trofimovich's generosity to allow us to access his research lab and relevant facilities at Concordia University.

FOREIGN LANGUAGE SPEECH LEARNING

Abstract

The current project longitudinally investigated the extent to which first-year Japanese university students developed their second language (L2) oral ability in relation to increased input in foreign language classrooms. Their spontaneous speech was elicited at the beginning, middle and end of one academic year, and then judged by linguistically trained coders for pronunciation, fluency, vocabulary and grammar qualities. According to the statistical analyses, the total amount of input (operationalized as number of English classes taken and L2 use outside of classrooms) was significantly related to the participants' quick and immediate development of fluency and lexicogrammar during the first semester. Their pronunciation development was mixed, either subject to continuous change over two academic semesters (for prosody) or limited within the timeframe of the study (for segmentals). Similar to naturalistic L2 speech learning, the findings support the multifaceted role of input in different areas of oral proficiency development in foreign language classrooms.

Key words: Second language speech, Foreign language education, Pronunciation, Fluency, Vocabulary, Grammar

FOREIGN LANGUAGE SPEECH LEARNING

The Role of Input in Second Language Oral Ability Development in Foreign Language Classrooms

Previous second language acquisition (SLA) research has extensively examined how adult second language (L2) learners improve their oral proficiency as they process an increasing amount of input in naturalistic settings while using the L2 on a daily basis (Flege, 2009). A growing number of researchers have begun to examine the extent to which such findings are generalizable to foreign language (FL) classroom conditions, which are by definition characterized by several hours of L2 input per week (e.g., Muñoz, 2008). In the context of 40 Japanese FL learners of English, the current study examined how FL education can longitudinally impact the segmental, prosodic, temporal, lexical and grammatical qualities of spontaneously produced L2 speech according to quantity (1 semester vs. 1 year) and type (form/content-based instruction, extra L2 use outside of classrooms) of input.

Background

Roles of Input in L2 Speech Learning

In the field of SLA, the relationship between quantity and quality of input, and successful language learning has been subject to extensive inquiry. According to Flege's (2009) oft-cited definition, input is referred to as "L2 vocal utterances the learner has heard and comprehended, including his own, regardless of whether these utterances have been produced correctly by L2 native speakers, or incorrectly by other non-native speakers of the L2" (p. 175). Under this broad definition, "input" is used as an umbrella term, as it subsumes any exposure to the target language during L2 comprehension as well as speech production. All existing theories have at least agreed with the idea that input is fundamental to SLA. A usage-based account, for instance,

FOREIGN LANGUAGE SPEECH LEARNING

posits that when L2 learners frequently encounter specific linguistic features in the input, they establish stronger connections between these features and the various contexts in which they occur. This style of associative learning develops effortless, fast and intuitive processing of the L2 as formulaic units in response to relevant situational and linguistic cues (N. Ellis, 2006).

According to the input-driven view on SLA, L2 development takes place as learners restructure and change their interlanguage representations in their brain via noticing and understanding input. Subsequently, this developing L2 system guides L2 learners to analyze/refine their own production in order to achieve enhanced output performance in the later stages of SLA, suggesting that input and output facilitate L2 development in a complementary fashion (Gass, 1997; VanPatten, 2002).

Previous studies using cross-sectional research designs have illustrated that the quality of L2 learners' speech is predicted by length of residence in an L2 speaking environment when the L2 is used as the main language of communication (but not the L1) (e.g., Flege & Liu, 2001). More recently, many scholars have longitudinally examined various dimensions of L2 learners' oral proficiency development after their arrival in an L2 speaking country. For example, much learning is likely to happen within the first three to four months of immersion (e.g., study-abroad) in terms of fluency (Segalowitz & Freed, 2004), and lexicogrammar accuracy (Mora & Valls-Ferrer, 2012) and complexity (Vercellotti, 2015). When it comes to pronunciation, learning trajectories appear to be complex. Whereas many L2 learners' phonological forms of word production quickly become intelligible (Munro & Derwing, 2008), the development of adequate and varied prosody (e.g., word stress, intonation) may require several years of L2 experience (Munro & Derwing, 2014). The acquisition of more accurate and refined segmental production,

FOREIGN LANGUAGE SPEECH LEARNING

in particular, may be a slow, gradual and extensive process that is thought to occur over long periods of time (e.g., 10+ years) (Saito, 2015; Flege, 2009; Munro & Derwing, 2013).

Taken together, the findings from cross-sectional and longitudinal studies have thus far shown that successful SLA in naturalistic settings can be achieved as a result of the dynamic interaction between input and linguistic domains, which in turn suggests a hierarchical difficulty inherent in adult L2 speech learning: fluency, lexicogrammar → prosody → segmental, from least to most difficult. In the current study, we took an exploratory approach towards examining the generalizability of this broad, admittedly simplified model of L2 speech learning model to another L2 learning context—foreign language classrooms.

Foreign Language Education

According to Muñoz (2008, p. 578-579), FL learning typically possesses the following features: (a) instruction is limited to 2-4 sessions of approximately 50 minutes per week; (b) exposure to the target language during these class periods may be limited in source (mainly the teacher), quantity (not all teachers use the target language as the language of communication in the classroom), and quality (there is great variability in teachers' oral fluency and general proficiency); (c) the target language is not the language of communication between peers; and (d) the target language is not spoken outside the classroom. L2 instruction in these FL contexts is typically notorious for its emphasis on decontextualized drill activities without sufficient conversational experience with native and other non-native speakers, which Larson-Hall (2008) referred to as the “minimal input” condition (p. 36). Furthermore, many foreign language students, such as Japanese English-as-a-Foreign-Language (EFL) learners (the focus of the study), have the tendency to study the target language for examination-driven purposes rather than communicatively-oriented purposes (Kozaki & Ross, 2008).

FOREIGN LANGUAGE SPEECH LEARNING

Although it is clear that few adult L2 learners can achieve native-like (or even highly advanced) proficiency solely based on FL instruction (Derwing & Munro, 2009), it is important, for several reasons, to further pursue the extent to which FL instruction alone can actually impact L2 speaking performance. First, a great number of L2 learners still learn the target language through explicit language-focused lessons in many FL settings (Schulz, 2001). DeKeyser (2007) suggested that what students have explicitly learnt in FL classrooms can determine the degree of success in terms of their future experiences overseas in business and academic settings, suggesting that the quality of foreign language education plays a key role in L2 learners' ultimate attainment in the long run. Second, while previous L2 education studies have constantly noted a significant role for instruction in adult SLA (for review, Ellis, 2012), these findings need to be interpreted with much caution, since many studies have selectively highlighted the effectiveness of a very brief amount of instruction on (exclusively) the development of L2 morphosyntax. For example, the mean length of the grammar teaching studies featured in Spada and Tomita (2010) approximately three hours, ranging from 20 minutes to nine hours.

Finally, a better understanding of the current state of FL education is crucial to designing and elaborating future educational reforms. In Japanese FL settings (the focus of the study), for example, the quality of learning/teaching materials has been gradually changing, as the Japanese Ministry of Education has made a continuous effort to update the FL syllabus with a strong focus on using English as a practical tool for communication (i.e., an Action Plan to Cultivate "Japanese with English Abilities"). An increasing number of Japanese FL students have thus begun to study English to pursue "short-term realistic goals related to examinations and grades" as well as "a somewhat vague long-term objective related to using English for international/intercultural communication" (Yashima, Zenuk-Nishide, & Shimizu, 2004, p. 121).

FOREIGN LANGUAGE SPEECH LEARNING

Re-examining the effectiveness of FL education on the development of L2 oral ability can generate valuable information for researchers, practitioners, and policy makers alike.

Several studies have been conducted to examine the role of input in FL settings. Muñoz (2006) examined how Spanish-Catalan bilingual learners of English with different starting ages (i.e., 8, 10, 14, 18+ years) improved on a range of linguistic abilities over eight years of FL instruction. The results of this study showed that adult learners noted a substantial gain from a smaller amount of L2 instruction than child learners did, although all groups' performances plateaued at points far below native-like proficiency. The study suggests that L2 instruction itself can be an efficient teaching method for adult (rather than child) learners due to their cognitive maturity (e.g., advanced logical and deductive reasoning), literacy knowledge (e.g., larger first language vocabulary size), and accumulated experience at school (e.g., familiarity with learning the L2 under foreign language classroom conditions) (Muñoz, 2008).

In the context of 56 Japanese EFL university students who had studied English for six years (Grades 7-12) through FL education alone without any experience abroad, Saito and Hanzawa (2016) conducted a cross-sectional investigation on how the participants' varied FL experience related to the segmental, prosodic and temporal aspects of their spontaneous speech. Results revealed that the students' success, especially in relatively difficult domains of L2 speech (segmentals and word stress) was strongly correlated with the number of hours they had spent studying L2 English inside and outside of classrooms.

Whereas these studies have generally supported the relative effectiveness of extensive FL instruction on adult L2 learners' *long-term* SLA, others have revealed the *limits* of L2 speech learning especially when instruction was delivered within a *short-term* time framework. For example, Muñoz and Llanes (2014) longitudinally examined the degree of foreign accent

FOREIGN LANGUAGE SPEECH LEARNING

(strongly tied to segmental accuracy) in two age groups—child vs. adult learners—with two different types of exposure—i.e., three months of formal instruction only vs. formal instruction plus study abroad. Their findings revealed that whereas both child and adult learners significantly reduced their foreign accentedness in the study-abroad context, no improvement was evident in the FL context (see also Mora & Valls-Ferrer, 2012).

It is clear that more research is required to further examine the role of input in driving L2 oral ability development in FL classrooms. Given that previous study-abroad research has clearly illustrated that L2 speech learning patterns greatly vary according to quantity and quality of input that learners actually process even during the same length of immersion in an L2 speaking environment (e.g., Segalowitz & Freed, 2004), there is a particular need for research focusing on the multifaceted characteristics of L2 input (the amount and type of input inside and outside of classrooms) and L2 oral ability (pronunciation, fluency, vocabulary and grammar). In this regard, the current study was designed to illustrate whether, to what degree, and how three different types of input specific to Japanese EFL settings—operationalized as the number of form- and content-based classes and extra L2 exposure outside of classrooms—were predictive of adult L2 learners' improvement in terms of segmental (vowels/consonants), prosodic (word stress, intonation), temporal (speech rate), lexical (appositeness, richness) and grammatical (accuracy, complexity) L2 speech performance over one academic year (two academic semesters lasting 30 weeks). Situated in a FL program in a Japan university, the speaking data of 40 first-year students was collected three time points: the beginning (T1) and end (T2) of the first semester, and the end of the second semester (T3).

FOREIGN LANGUAGE SPEECH LEARNING

Research question and predictions. In light of the learning difficulty hierarchy suggested by previous naturalistic SLA research (fluency, lexicogrammar < prosody < segmentals), the following research question and three predictions were formulated:

RQ: How do Japanese university students differentially improve their oral abilities in relation to quantity and quality of L2 input in FL classroom settings?

1. The Japanese EFL students' language may become fluent (delivered at an optimal speech rate without too many pauses), and their lexicogrammar use may become accurate and complex, especially during the first term.
2. These students may gradually enhance prosodic accuracy especially during the second term.
3. Their improvement in segmental accuracy may be limited under FL conditions.

Method

Participants

In total, 40 first-year students majoring in a wide variety of social sciences and humanities (e.g., business, economics, psychology, and international relations) at a large university in Japan participated in the study (their age = 18-19 years). According to the initial language background questionnaire completed at T1, they had started learning English during secondary school. Between Grades 7 and 12 (6 years), they had taken form-focused English lessons as well as attended prep schools (a total of several hours per week) in order to develop their awareness of cross-cultural communication as well as prepare for their university entrance

FOREIGN LANGUAGE SPEECH LEARNING

exams. According to the national syllabus (i.e., an Action Plan to Cultivate “Japanese with English Abilities”), all Japanese FL students are required to enroll in one-to-two hours of oral communication classes per week (either taught by native or non-native teachers). However, it has been often pointed out that these language lessons tend to exclusively concern grammar translation without sufficient opportunities to use language for communicative purposes (for the “realities” of FL classrooms in Japan, see Nishino & Watanabe, 2008). The participants in the study also reported the lack of any prior experience in staying and studying abroad nor in taking any subject lessons in English prior to the project.

Content of FL Input

Following previous classroom SLA research, L2 input in the current study was operationalized according to three different dimensions which have been found to be associated with successful language learning in FL classroom settings: (a) length of instruction (the total hours of EFL classes) (Munoz, 2008), (b) type of instruction (form- vs. meaning-oriented classes) (Spada & Tomita, 2010), and (c) extra L2 use outside of classrooms (Munoz, 2014).

At the end of the first and second semesters, the participating students filled in a questionnaire about the total number of English classes that they had enrolled in as well as the amount of time they had spent practicing English through informal conversations with other native and non-native speakers of English (international students at the university). Next, we individually interviewed the participants in order to determine what types of FL instruction they had actually received. During the interview, they were asked to self-report the details every class they had taken in English in relation to (a) class size (the number of students per class); (b) instructor (natives, non-natives); (c) content of class (e.g., form- vs. meaning-oriented, the ratio of lectures vs. in-class discussion, assessment methods); (d) understanding level (the degree of

FOREIGN LANGUAGE SPEECH LEARNING

understanding of L2 English in class on a 7-point scale); (e) output level (the frequency of L2 speaking in class on a 7-point scale); and (f) interlocutor types (with whom [natives vs. non-natives vs. Japanese] did the participants use L2 English in class). For the interview note we used, please find Appendix.

In conjunction with the assumed variability in the instructional content, size and teaching styles of their FL classes, we adopted a context-specific classification style—form-focused vs. content-based. Our intuitive, straightforward approach differs from more detailed, restricted coding schemes that researchers typically use for more controlled and brief instruction, such as “explicit vs. implicit” (Rebuschat, 2013) and “input- vs. output-based” (Shintani, Ellis, & Li, 2013) distinctions. Table 1 provides the descriptive statistics of the participants’ EFL experience inside (the number of form-focused and content-based classes) and outside (extra L2 use) of classrooms during the project.

In keeping with Spada (1997), form-focused instruction was defined as “any pedagogical effort which is used to draw the learners’ attention to language form either implicitly or explicitly” (p. 73). According to our casual classroom observations¹ and analyses of the syllabi at the university, the content of these form-focused lessons was designed to develop various L2 English skills by offering a wide range of practice opportunities spanning reading, listening, writing and speaking activities. These classes were taught by native, non-native, and Japanese teachers alike. Importantly, the final grades for these classes were assessed based on L2 knowledge (e.g., pronunciation, vocabulary and grammar knowledge), L2 proficiency (reading, listening and writing abilities), or their mastery of academic L2 skills (e.g., the organization and clarity of discussion in L2 academic writing on a given topic). Whereas all the participants were

FOREIGN LANGUAGE SPEECH LEARNING

registered in at least one form-focused lesson per week in accordance with university requirements, the number of classes they were actually taking widely varied.

As defined by Richards and Rodgers (2001), content-based instruction is “an approach to second language teaching in which teaching is organized around the content or information that students will acquire, rather than around a linguistic or other type of syllabus” (p. 204).

According to the results of the individual interviews, the participants were registered in a wide range of humanities and social sciences courses, such as business, marketing, politics, economics, international relations, communication studies, psychology, history, film, and music. All classes were taught in L2 English without any use of L1 Japanese, irrespective of the L1 status of the teachers. Unlike form-focused lessons, the primary focus of their content-based lessons was understanding the subject matter, with the final grades being based on the participants’ academic (rather than linguistic) knowledge and performance. Since content-based instruction in L2 English was optional at the university, it could be considered as a proactive way for students to increase the amount of L2 input they received. The amount of content-based instruction received by the participants also varied greatly (0-90 hours per semester).

TABLE 1

Speaking Task

It has been shown that adult L2 learners can carefully monitor their production by drawing on their explicit linguistic knowledge, especially when their speech is elicited via controlled tasks (e.g., word and sentence reading) (Spada & Tomita, 2010). To tap into adult L2 learners’ oral performance at a spontaneous level, many instructed SLA researchers have emphasized the importance of adopting free-constructed speech measures (e.g., picture narratives).

FOREIGN LANGUAGE SPEECH LEARNING

Many L2 speech researchers have used an eight-frame cartoon to elicit relatively long spontaneous speech samples (1-3 minutes) (e.g., Derwing & Munro, 1997). Building on this line of thought, in our previous project which worked specifically with Japanese learners of English (Saito, 2015; Saito, Trofimovich, & Isaacs, 2016a), we elaborated a timed picture description task to allow not only experienced, but also inexperienced L2 learners to produce a certain length of semi-structured speech without too many pauses and repetitions. In this task, the participants explained seven pictures, with three key words given as cues. To encourage the participants to speak under time pressure, they were provided with only five seconds of planning time prior to each picture description. The first four (out of seven) pictures were used for practice to familiarize the students with the task procedure; the remaining three pictures were used for the final analyses. It needs to be emphasized that the nature of the speech samples in this study could be considered as “semi” spontaneous, as the participants were able to use three key words (instead of looking for all lexical items on their own)—a methodological limitation we will revisit in the Conclusion section.

Materials. The three target pictures described (a) a table left alone in a driveway in a rain storm (keywords: rain, table, driveway); (b) three guys playing rock music with one lead vocalist and two guitar players (keywords: three guys, guitar, rock music); and (c) a long straight road with some clouds in the blue sky (keywords: blue sky, road, cloud). The keywords were intentionally chosen to guide Japanese learners to use problematic segmental and syllabic features, such as the neutralization of the English /r/-/l/ contrast (“rain, rock, brew, crowd” vs. “lane, lock, blue, cloud”), and the insertion of extra epenthetic vowels between consecutive consonants (/dɚraɪvə/ for “drive,” /θəri/ for “three,” /səkaɪ/ for “sky”) and after word-final consonants (/teɪbələ/ for “table,” /myuzɪkə/ for “music”) in borrowed words (i.e., Katakana).

FOREIGN LANGUAGE SPEECH LEARNING

Procedure. All speech samples were recorded in a soundproof booth at the Japanese university using a Marantz PMD 660 recorder with a Shure SM 10A-CN microphone. We adopted a 44.1 kHz sampling rate with 16-bit quantization. All instructions were delivered in Japanese by the researcher (a native speaker of Japanese) so that all participants fully understood how to complete the task. After the first four practice pictures were presented in a randomized order, the three target pictures were randomly shown to the participants.

We selected -the first 10 sec of each picture description, and combined and stored them in one single WAV file per talker (30 sec per talker). To keep the samples as natural as possible, we carefully edited and cut full phrases, which resulted in some variance in the length of the speech samples. The average length of the samples was 30.5 sec ranging from 25.6 to 42.5 sec. The mean number of words for each file was 40.1 words, ranging from 21-58 words. In keeping with previous L2 speech studies (e.g., Derwing & Munro, 1997 for 15-30 sec; Isaacs & Trofimovich, 2012 for 30 sec), the length of the files was considered to provide sufficient linguistic information for phonological, temporal, lexical and grammatical analyses (Isaacs & Trofimovich, 2012). In total, 120 speech samples were generated (40 Japanese students at T1, T2, and T3).

As clearly noted in other similar studies where human raters are asked to listen to and evaluate a number of L2 speech samples (e.g., Derwing & Munro, 1997), using relatively short speech samples for each participant (30sec) is a crucial methodological decision in order to minimize the amount of listener fatigue—a relatively strong confounding variable for L2 speech assessment (the longer rating sessions last, the more fatigue listeners feel, the less consistent/reliable their rating behaviours become). In addition, it has been shown that extensive

FOREIGN LANGUAGE SPEECH LEARNING

exposure to L2 speech samples could significantly alter raters' judgement patterns (e.g., Flege & Fletcher, 1992).²

Notably, the same materials were used at the three different testing points (T1, T2 and T3), as we prioritized eliciting comparable speech samples from the participants over one academic year (for similar methodological decisions in longitudinal L2 speech studies, see Munro & Derwing, 2008, 2013). Although we acknowledge the possibility of “practice” effects (due to taking the same test more than twice), we assume that such test-retest gains were minimized in the current study. First, the testing intervals were relatively long (15 weeks). Second, the participants performed the first four picture descriptions as “practice” during each testing session. As such, their task familiarity was reduced before they worked on the final three photos (that were used for linguistic analyses). Third, as shown in the Results section (presented below), the participants did not significantly improve on most of the linguistic domains of their L2 speech over the first semester (T1 → T2), suggesting that simply taking the same test twice did not impact the longitudinal development of the FL students' L2 oral skills (the results rather pointed out that their complex improvement patterns were substantially influenced by the amount and type of L2 input they had processed in FL classrooms).

Pronunciation, Fluency, Vocabulary and Grammar Analyses

In the field of SLA, L2 oral ability has been conceptualized as a composite construct spanning various linguistic domains such as pronunciation, fluency, vocabulary, and grammar, with the assumption that L2 learners differentially enhance the phonological, temporal, lexical, and grammatical qualities of their L2 speech—in other words, it represents a componential view of L2 speech learning (e.g., De Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2012). A number of L2 speech researchers have drawn on objective measures (e.g., acoustic analyses) to investigate

FOREIGN LANGUAGE SPEECH LEARNING

various linguistic phenomena. However, the use of these measures may be limited to *controlled* speech samples, whereby researchers can pinpoint target features independent of the influence of surrounding phonetic and lexical contexts (Piske, MacKay, & Flege, 2001). When it comes to the linguistic analysis of *spontaneous* speech samples, L2 speech researchers tend to use trained raters' expert evaluations of segmentals (Piske, Flege, MacKay, & Meador, 2011), word stress (Field, 2005), intonation (Derwing & Munro, 1997), fluency (Bosker, Pinget, Quené, Sanders, & de Jong, 2013), and vocabulary (Crossley, Salsbury, & McNamara, 2015). In this subjective judgement paradigm, highly experienced raters (e.g., linguists, ESL/EFL professionals) are carefully selected and trained to judge specific aspects of target language features by referring to detailed descriptors while listening to spontaneous speech samples.

In the current study, we adopted the expert judgement approach that was previously elaborated and validated in precursor research (Saito, Trofimovich, & Isaacs, 2016b). The trained raters first listened to audio samples to assess four phonological and temporal categories—(a) segmentals (substitution, omission, or insertion of individual consonants or vowels); (b) word stress (misplaced or missing primary stress); (c) intonation (appropriate/varied versus incorrect and monotonous use of pitch); and (d) speech rate (speed of utterance delivery). After the audio recordings were transcribed with all orthographic markings of pausing (e.g., uh, um, oh, ehh) removed, the raters read the final transcripts in order to assess four lexicogrammatical categories: (a) lexical appropriateness (accuracy of vocabulary); (b) lexical richness (varied and sophisticated use of vocabulary); (c) grammatical accuracy (errors in word order, grammar endings, agreement); and (d) grammatical complexity (use of sophisticated, non-basic grammar).

Expert raters. Following Isaacs and Thomson (2013), five expert raters (3 males, 2 females) were recruited based on their extensive linguistic and pedagogical experience. All of

FOREIGN LANGUAGE SPEECH LEARNING

them were graduate students in Applied Linguistics at the Canadian university. They reported that they had received training in phonetics and phonology. Additionally, they reported a sufficient amount of teaching experience in ESL/EFL settings ($M = 4.0$ years from 2 to 6 years). According to the language background questionnaire, they demonstrated relatively high familiarity with Japanese-accented English ($M = 4.4$ from 1 = *Not at all* to 6 = *Very much*). None of the raters reported any hearing problems.

Procedures. To rate the spontaneous speech samples, the raters engaged in a two-step procedure. First, they listened to each audio file randomly played via the MATLAB interface. Immediately after listening to each file, the raters used a moving slider (1000-point scale: 1 = *non-targetlike*, 1000 = *targetlike*) to simultaneously evaluate the four rating categories (segmentals, word stress, intonation, speech rate). To ensure the complex nature of the phonological analyses (rating four pronunciation categories—segmentals, word stress, intonation, speech rate—while listening to one speech sample), the raters had the option to replay each file multiple times until they felt satisfied with their judgements. For the transcript-based measures, the raters read each speech transcript displayed on a computer screen and simultaneously used the same moving slider to evaluate the text on four categories (lexical appropriateness and richness, grammatical accuracy and complexity). Similar to the audio-based measures, they were asked to spend a sufficient amount of time analyzing each file. For on-screen labels, see Supporting Document. The rating sessions took place in a quiet room on three different days: Day 1 (about 3 hours) for training, Day 2 (3 hours) for the audio-based ratings, and Day 3 (1.5 hour) for the transcript-based ratings.

Training sessions (Day 1). The raters first received thorough instructions from a trained research assistant on the eight domains of pronunciation, fluency, vocabulary, and grammar (for

FOREIGN LANGUAGE SPEECH LEARNING

training scripts, see Supporting Document). They practiced the rating procedure with 40 speech samples of non-native speakers with varied proficiency levels—the dataset derived from Isaacs and Trofimovich (2012)—first for the audio-based measures, and second for the transcript-based measures. Subsequently, their subjective rating scores were compared with the actual linguistic properties of the tokens (reported in Isaacs & Trofimovich, 2012). As summarized in Table 2, the raters' pronunciation, fluency, vocabulary and grammar scores were significantly correlated with certain (but not all) objective measures.³

TABLE 2

Rating sessions (Days 2 & 3). After receiving recapped instructions on the four pronunciation/fluency categories from a trained assistant, the raters familiarized themselves with the picture prompts and key words. They rated five audio samples of Japanese learners (not included in the main analysis) as practice. For each sample, the raters were asked to explain their decisions on the four categories, and received feedback from the trained assistant on the accuracy of their analyses. They then proceeded to rate the 120 speech samples (40 talkers × three testing points). They listened to the samples in a randomized order.

The next day, the raters received recapped instructions on the four vocabulary/grammar categories, practiced the procedure with the five trial samples, and then moved to assessing 120 written files which were presented in a randomized order via the MATLAB interface. At the end of the final session, the raters were asked to self-report their understanding of each category (1 = *I did not understand at all*, 9 = *I understand this concept well*) and comfort in using each (1 = *very difficult*, 9 = *very easy and comfortable*). They showed relatively high level of understanding ($M = 7.8-9.0$) and comfort ($M = 7.2-8.5$) for all linguistic categories.

Results

FOREIGN LANGUAGE SPEECH LEARNING

Group Improvement

The first aim of the statistical analyses was to examine the extent to which the FL students as a group improved their oral ability over the first (T1→T2) and second semesters (T2→ T3). According to the results of one-sample Kolmogorov-Smirnov tests, the dataset was normally distributed according to different linguistic domains (pronunciation, fluency, lexicogrammar) at each testing point (T1, T2, T3) ($p > .05$). Nevertheless, due to small sample size ($N = 40$), we chose to conduct more conservative statistical analyses—nonparametric Wilcoxon Signed Ranks tests—to examine the presence/absence of significant changes over the first (T1→T2) and second (T2→T3) semesters with alpha set to .025 (Bonferroni corrected).

As summarized in Table 3, the participating students significantly improved only their speech rate after receiving one semester of FL instruction ($p < .001$) with a medium effect size ($d = 0.41$). Their improvement over the second semester was only marginally significant ($p = .033$) with a medium effect size ($d = 0.69$). The lack of statistical significance in the other domains of L2 speech (segmentals, word stress, intonation, lexicogrammar) indicated a great deal of individual variance among the students, arguably due to their different EFL experience in quantity (e.g., certain students processed more input each semester) and quality (e.g., certain students received input from various resources inside and outside of classrooms) (see Table 1).

TABLE 3

Quantity of Input and L2 Speech Learning

The second aim of the analysis was to examine whether and to what extent the varied amount of instructional gains among the participating students could be related to the total amount of their received L2 input inside and outside of classrooms for the first (range: 45-223 hours) and second semester (range: 33-220 hours) (see Table 1). In order to isolate the influence

FOREIGN LANGUAGE SPEECH LEARNING

of the participants' previous EFL experience (the amount of L2 input during six years of EFL experience prior to the project; see Saito & Hanzawa, 2016), a set of partial correlation analyses were conducted with their gain scores over the first semester (T1→T2) and second semester (T2→T3) as dependent variables, and their total hours of L2 use inside and outside of classrooms as independent variables, while controlling for their proficiency scores at T1 and T2, respectively. The alpha value was set to .05 for this analysis.

As shown in Table 4, the input factor was significantly predictive of the extent to which the participants improved their prosodic, temporal, lexical and grammatical aspects of L2 speech for the first semester. While there was no significant role of input in the further development of the participants' fluency (speech rate) and lexicogrammar (appropriateness, richness, complexity) performance after the first term, the input factor continued to significantly and marginally predict the participants' prosodic improvement—word stress ($p = .04$) and intonation ($p = .06$)—for the second semester. In contrast, no significant relationship between input and segmental improvement was found throughout the project.

TABLE 4

Type of Input and L2 Speech Learning

The final aim of the analyses was to examine whether and how three different types of L2 input specific to the context of the study—form-focused and content-based classes and extra L2 use outside of classrooms—differentially related to the participants' gain scores during the first (T1→T2) and second semesters (T2→T3). Similar partial correlation analyses were performed to examine the associations between three different input factors (a total hours of form/content classes and L2 use outside of classrooms) and the participants' gain scores with their initial

FOREIGN LANGUAGE SPEECH LEARNING

proficiency scores factored out. The alpha value was corrected to a $p < .012$ level for this analysis.

As summarized in Table 5, the total number of content-based classes was significantly predictive of the prosodic (word stress, intonation) qualities of the participants' L2 speech performance over the first and second semesters. The predictive power of the number of content-based classes was also significant for the participants' vocabulary (appropriateness, richness) and grammar (accuracy, complexity) over for the first (but not the second) semester. The number of form-focused classes was significantly associated with the participants' prosodic (word stress, intonation) performance ($p < .012$) for the first semester. Finally, no significant input and proficiency relationship was found in any contexts related to the participants' segmental performance and the amount of extra L2 use outside of classrooms.

TABLE 5

Discussion

In light of the growing interest in the generalizability of findings from naturalistic SLA contexts to FL classrooms (Muñoz, 2008), the current study was a first attempt to longitudinally examine the extent to which L2 learners can improve on various constructs of L2 speech—segmentals, prosody, fluency, vocabulary and grammar—under FL classroom conditions over one academic year. In answering the main research question (i.e., which L2 oral abilities are relatively amenable to one year of FL instruction), the current study provided three crucial findings. First, the participants' improvement across the various linguistic domains was significantly correlated with the total number of hours they had spent using L2 English during the first term (45-223 hours). Second, these students continued to improve their prosodic accuracy (word stress, intonation) in relation to increased L2 input over the second term (33-220

FOREIGN LANGUAGE SPEECH LEARNING

hours). Third, their segmental performance was resistant to change under FL classroom conditions regardless of the amount of L2 input over one academic year (78-444 hours).

In line with previous longitudinal findings in naturalistic contexts (e.g., Flege, 2009; Saito, 2015), and our predictions mentioned earlier (Saito & Hanzawa, 2016), the results presented here suggest that L2 input plays a crucial role in classroom L2 speech learning with similar learning patterns according to linguistic domain (more fluent, appropriate and complex use of lexicogrammar → more refined and accurate pronunciation). As shown in the current study, the input-proficiency link was particularly strong during the initial stage of fluency and lexicogrammar development, when the first-year university students were spending the first semester under new university-level FL classroom conditions. The results are consistent with similar findings on L2 fluency (e.g., Segalowitz & Freed, 2004) and lexicogrammar (Mora & Valls-Ferrer, 2012) development during short-term study-abroad. In contrast, the students' pronunciation learning was mixed, either occurring over one academic year (for word stress, intonation) or being limited (for segmentals). This is perhaps not surprising, as it is claimed that pronunciation development requires much exposure to L2 input (Flege, 2009) and/or certain optimal individual difference profiles, such as early age of learning (Abrahamsson & Hyltenstam, 2009), high-level aptitude (Skehan, 2015) and professional motivation (Moyer, 1999). The limited effect of short-term instruction on pronunciation learning in FL classrooms is also consistent with previous FL studies (Muñoz & Llanes, 2014).

As for type of L2 input, the current study pointed out that the students' gain scores were specifically tied to the number of content-based classes they were taking. Given that enrollment in these classes was optional at the university (summarized in Table 1), the results suggest that certain students could attain successful L2 speech learning in FL contexts by intentionally

FOREIGN LANGUAGE SPEECH LEARNING

increasing the amount of L2 input they receive. Our interpretation roughly echoes what instructed SLA research generally emphasizes for successful foreign language learning—a well-balanced focus on form *and* meaning during L2 instruction (Spada & Tomita, 2010). At the same time, however, any interpretations as to the optimal type of input for successful classroom L2 speech learning should be considered as tentative given the exploratory nature of the method we adopted: We quantified the nature of L2 input based on the analysis of the interview data at the end of the first and second semesters by using an admittedly simplistic coding scheme for different types of L2 instruction (i.e., form-focused vs. content-based instruction). For a similar methodological reason, we are also hesitant to make any generalizations in regards to the non-significant predictive power of extra L2 use for any domain of L2 speech. Although this factor was simply surveyed based on the number of hours the participants spent on informal conversation activities with other non-native and native international students at the university, it remained unclear whether such retrospective method (self-reports) could reliably provide a detailed picture of the participants' EFL experience over one academic year.

One promising research direction could be to conduct longitudinal research in EFL contexts while controlling for the type of input as one “independent” variable. For example, Trofimovich, Lightbown, Halter and Song (2009) investigated the role of two different types of instruction—comprehension-based instruction (extensive reading) vs. production-based instruction (audiolingual training)—in the development of young EFL learners' L2 oral abilities over the course of two years. Whereas the production-based instruction group continued to improve their overall comprehensibility and fluency throughout the project, the comprehension-based instruction group's gains plateaued after the first year. Saito and Akiyama (2016) recently examined the extent to which Japanese EFL university students improved their L2 oral

FOREIGN LANGUAGE SPEECH LEARNING

proficiency over one academic semester as a result of two different types of instructional treatments. Whereas the control group (which was involved with lexicogrammar exercise activities) did not show any significant improvement in their oral abilities, a great deal of successful L2 speech learning was observed among the experimental group (which participated in weekly, dyadic task-based interaction activities with native interlocutors in the US). In sum, these two longitudinal studies (Saito & Akiyama, 2016; Trofimovich et al., 2009) have supported the importance of production and interaction opportunities in order to improve L2 speech in an efficient and effective manner in EFL classroom settings.

Conclusion

The current study investigated the role of input in the longitudinal development of 40 first-year university students' oral ability in EFL classrooms over one academic year. The findings concurred with previous longitudinal studies on naturalistic L2 speech learning in that the students showed quick improvement in fluency and lexicogrammar within first semester; and their pronunciation development was subject to continuous change over two academic semesters (for prosody) or limited within the time framework of the study (for segmentals).

Due to the exploratory nature of the project, we would like to close this paper by addressing several limitations for the purpose of providing a more comprehensive picture of the pedagogical potential/limits of FL education for L2 speech learning.

Our first obvious limitation was the small sample size ($N = 40$), due to the longitudinal nature of the project (which required some participants to visit the researcher's office three times over one year). Given that attrition can be a challenge for any longitudinal research in the field of SLA, our findings need to be replicated with larger samples. In addition, we acknowledge that the nature of the project was descriptive rather than experimental, and lacked any elaboration or

FOREIGN LANGUAGE SPEECH LEARNING

control of the quality and quantity of instructional treatment in a proactive manner (mainly because this study was designed to evaluate the pedagogical potential/limits of the existing FL program). To establish a stronger link between what students have learned and how much they have improved, future researchers need to conduct such experimental studies, especially with a long-term perspective (cf. Trofimovich et al., 2009).

Next, several limitations need to be acknowledged in terms of the construct validity of the speaking task. The findings in the study were exclusively based on a single task (i.e., timed picture description) with a relatively short length of speech samples (i.e., 30 sec). Given that the participants could rely on three key words for each picture description, the nature of the speech samples can be considered as “semi” spontaneous at best. Future studies first need to elicit “extemporaneous” speech which simulates speech in real-life situations, where L2 learners use the target language for communicative purposes.

Importantly, previous L2 research has shown that the L2 learners’ linguistic accuracy, fluency, and complexity of speech can be greatly influenced by certain task modalities. It would be intriguing if future studies adopt a range of speaking tasks with and without pre-task and online planning time (Yuan & Ellis, 2003), task repetition (Ahmadian & Tavakoli, 2011), and interaction (Derwing, Rossiter, Munro, & Thomson, 2004). Such tasks can also be repeated with an increasing time pressure—another variable influencing L2 speech performance (Thai & Boers, 2016).

Whereas we ensured the validity of the speech sample length (30 sec) following the research standards in L2 speech research (e.g., Derwing & Munro, 2009), the short length of the speech samples ($M = 40.1$ words) did not allow us to examine how FL students improved their vocabulary and grammar aspects of L2 speech—the linguistic domains strongly tied with L2

FOREIGN LANGUAGE SPEECH LEARNING

compressibility (Isaacs & Trofimovich, 2012). In this regard, future research needs to elicit longer speech samples (e.g., Yuan & Ellis, 2003 for 3 min) in order to incorporate a range of robust lexical analyses which require oral texts consisting of 100-200 words (Koizumi & In'nami, 2012).

Endnotes

1. Efforts were made to visit as many form-focused classes enrolled in by the participants as possible (at least once throughout the project). During individual interviews, we also asked the participants to self-report on the details of each form-focused and content-based class that they had taken (see Appendix). However, it needs to be emphasized that we did not aim to use such information to further code/categorize different types of instructional activities (e.g., the ratio of listening, reading, writing and speaking activities), nor did we aim to generalize any type-of-instruction pattern beyond the broad categories/terms used in the study—form-focused instruction. This is because doing so is extremely difficult, especially for L2 longitudinal research of this kind (e.g., the participants belonged to different classes across different programs, and the content of each class inevitably changed over a period of one academic year).

2. The subjective judgement approach is essentially different from the objective judgement approach, where longer speech samples (1-3 min) are analyzed based on a range of acoustic and corpus measures. While the latter approach is appropriate for assessing controlled speech, the former approach has been used for analyzing spontaneous speech.

3. The details of the training sessions were documented in our prior research (Saito et al., 2016b).

References

- Abrahamsson, N. & Hyltenstam, K. (2009). Age of acquisition and nativelikeness in a second language – listener perception vs. linguistic scrutiny. *Language Learning*, 59, 249–306.
- Ahmadian, M. J., & Tavakoli, M. (2011). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research*, 15, 35-59.
- Bosker, H. R., Pinget, A.-F., Quené, H., Sanders, T., & De Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, 30, 159-175.
- Crossley, S. A., Salsbury, T., & Mcnamara, D. S. (2015). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics*, 36, 570-590.
- De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition*, 34, 5–34.
- DeKeyser, R. (Ed.). (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge, UK: Cambridge University Press.
- Derwing, T.M. (2003). What do ESL students say about their accents? *Canadian Modern Language Review*, 59, 545-564.
- Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 12, 1–16.
- Derwing, T. M. & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, 42, 476-490.
- Derwing, T.M., Rossiter, M.J., Munro, M.J. & Thomson, R.I. (2004). L2 fluency: Judgments on different tasks. *Language Learning*, 54, 655-679.
- Ellis, R. (2012). *Language teaching research and language pedagogy*. Chichester, UK: Wiley-Blackwell.
- Flege, J. E. (2009). Give input a chance! In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp.175-190). Clevedon: Multilingual Matters.
- Flege, J., & Fletcher, K. (1992). Talker and listener effects on the perception of degree of foreign accent. *Journal of the Acoustical Society of America*, 91, 370-389.
- Flege, J. & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, 23, 527-552.
- Gass, S. (1997). *Input, interaction, and the second language learner*. Mahwah, NJ: Erlbaum.
- Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly*, 10, 135-159.
- Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition*, 34, 475-505.
- Koizumi, R., & In'nami, Y. (2012). Effects of text length on lexical diversity measures: Using short texts with less than 200 tokens. *System*, 40, 554-564.
- Kozaki, Y., & Ross, S. J. (2011). Contextual dynamics in foreign language learning motivation. *Language Learning*, 61, 1328-1354.
- Lambert, C., & Kormos, J. (2014). Complexity, Accuracy, and Fluency in Task-based L2 Research: Toward More Developmentally Based Measures of Second Language Acquisition. *Applied Linguistics*, amu047.

FOREIGN LANGUAGE SPEECH LEARNING

- Larson-Hall, J. (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research*, 24, 35-63.
- Mora, J. C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly*, 46, 610-641.
- Moyer, A. (1999). Ultimate attainment in L2 phonology: The critical factors of age motivation and instruction. *Studies in Second Language Acquisition*, 21, 81-108.
- Muñoz, C. (2006). The effects of age on foreign language learning: The BAF Project. In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 1-40). Clevedon: Multilingual Matters.
- Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics*, 24, 578-596.
- Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics*, 35, 463-482.
- Muñoz, C. & Llanes, A. (2014). Study abroad and changes in degree of foreign accent in children and adults. *The Modern Language Journal*. 98, 432-449.
- Munro, M. & Derwing, T. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning*, 58, 479-502.
- Munro, M. J., & Derwing, T. M. (2013). English L2 vowel acquisition over seven years. In J. Levis & K. LeVelle (Eds.). *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference*. Ames, IA: Iowa State University.
- Munro, M., & Derwing, T. (2014, March). *How different can you get? Ten-year learning trajectories for two L2 groups*. Paper presented at the annual conference of the American Association for Applied Linguistics, Portland, OR.
- Piske, T., MacKay, I., & Flege, J. (2001). Factors affecting degree of foreign accents in an L2: a review. *Journal of Phonetics*, 29, 191-215.
- Piske, T., Flege, J., MacKay, & Meador, D. (2011). Investigating native and non-native vowels produced in conversational speech. In M. Wrembel, M. Kul & Dziubalska-Kołodziejczyk, K. (Eds.), *Achievements and perspectives in the acquisition of second language speech: New Sounds 2010* (pp. 195-205). Switzerland: Peter Lang.
- Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, 63, 595-626.
- Richards, J. & Rodgers, T. (2001). *Approaches and Methods in Language Teaching, Second Edition*. Cambridge: Cambridge University Press.
- Schulz, R. A. (2001). Cultural differences in student and teacher perceptions concerning the role of grammar instruction and corrective feedback: USA-Colombia. *The Modern Language Journal*, 85, 244-258.
- Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, 26, 173-199.
- Shintani, N., Li, S., & Ellis, R. (2013). Comprehension-based versus production-based grammar instruction: A meta-analysis of comparative studies. *Language Learning*, 63, 296-329.
- Skehan, P. (2015). Foreign language aptitude and its relationship with grammar: A critical overview. *Applied Linguistics*, 36, 367-384.
- Saito, K. (2015). Experience effects on the development of late second language learners' oral proficiency. *Language Learning*, 65, 563-595.

FOREIGN LANGUAGE SPEECH LEARNING

- Saito, K., & Akiyama, Y. (2016). Video-based interaction, negotiation for comprehensibility, and second speech learning: A longitudinal study. *Language Learning*. doi: 10.1111/lang.12184
- Saito, K., & Hanzawa, K. (2016). Developing second language speaking proficiency in foreign language classrooms: The role of the length and focus of instruction and individual differences. *Applied Psycholinguistics*, 37, 813-840.
- Saito, K., Trofimovich, P., & Isaacs, T. (2016a). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*, 37, 217-240.
- Saito, K., Trofimovich, P., & Isaacs, T. (2016b). Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*. doi: 10.1093/applin/amv047.
- Spada, N. (1997). Form-focused instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching*, 29, 73-87.
- Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, 60, 263-308.
- Thai, C., & Boers, F. (2016). Repeating a monologue under increasing time pressure: Effects on fluency, complexity, and accuracy. *TESOL Quarterly*. doi: 10.1002/tesq.232
- Trofimovich, P., Lightbown, P. M., Halter, R., & Song, H. (2009). Comprehension-based practice: The development of L2 pronunciation in a listening and reading program. *Studies in Second Language Acquisition*, 31, 609-639.
- VanPatten, B. (2002). Processing instruction: An update. *Language Learning*, 52, 755-803.
- Vercellotti, M. L. (2015). The development of complexity, accuracy, and fluency in second language performance: A longitudinal study. *Applied Linguistic*. doi: 10.1093/applin/amv002
- Yashima, T., Zenuk-Nishide, L., & Shimizu, K. (2004). The influence of attitudes and affect on willingness to communicate and second language communication. *Language Learning*, 54, 119-152.
- Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, 24, 1-27.

FOREIGN LANGUAGE SPEECH LEARNING

Table 1

Summary of the Participants' EFL Experience during the Project (1st Semester, 2nd Semester)

	Total hours of input (inside and outside of classrooms)			Total hours of form-oriented classes (hr)			Total hours of content-based classes (hr)			Total hours of extra English use (hr)		
	<i>M</i>	<i>SD</i>	<i>Range</i>	<i>M</i>	<i>SD</i>	<i>Range</i>	<i>M</i>	<i>SD</i>	<i>Range</i>	<i>M</i>	<i>SD</i>	<i>Range</i>
Semester 1	119.9	60.5	45-223	79.0	22.6	45-120	38.0	39.3	0-90	3.0	7.1	0-35
Semester 2	96.2	54.9	33-220	58.0	22.2	16.5-79.5	31.3	34.0	0-99	6.9	12.2	0-42
Total (Terms 1 + 2)	216.1	113.3	78-444	95.2	16.6	52.5-120	69.3	72.0	0-189	9.9	17.7	0-77

FOREIGN LANGUAGE SPEECH LEARNING

Table 2

Correlations between Pronunciation, Fluency, Vocabulary and Grammar Ratings and Coded Linguistic Variables from Isaacs and Trofimovich (2012)

Rated Variable	Linguistic dimensions	Correlations
<u>A. Pronunciation</u>		
Segmentals	Vowel/consonant errors	.64*
Word stress	Word stress errors	.72*
Intonation	Intonation errors	.54*
Rhythm	Vowel reduction	.74*
<u>B. Fluency</u>		
Speech rate	Articulate rate	.43*
	Mean Length of Run	.79*
	No. of filled pause	.41
	No. of unfilled pause	.49*
<u>C. Vocabulary</u>		
Appropriateness	Lexical errors	.50*
Richness	Token frequency	.69*
	Type frequency	.74*
	Infrequent word ratio	.23
<u>D. Grammar</u>		
Accuracy	Word order and morphology errors	.67*
Complexity	Subordination clause ratio	.44*

Note. * $\alpha < .05$

FOREIGN LANGUAGE SPEECH LEARNING

Table 3

Results of the Production Test Scores over Time (1000-point: 0 = Non-targetlike, 1000 = Targetlike)

	T1		T2		T3		1st semester (T1→T2)			2 nd semester (T2→T3)		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>z</i>	<i>p</i>	<i>d</i>	<i>z</i>	<i>p</i>	<i>d</i>
Segmentals	391	103	388	101	397	84	-.309	.757	0.02	-.773	.440	0.06
Word stress	442	98	440	90	466	89	-.067	.946	0.02	-1.72	.085	0.25
Intonation	397	138	430	158	447	146	-1.94	.052	0.22	-.323	.747	0.35
Speech rate	446	182	520	177	572	178	-3.58	<.001*	0.41	-2.13	.033	0.69
Lexical appropriateness	760	90	759	82	780	74	-.282	.778	0.05	-1.47	.139	0.24
Lexical richness	394	116	400	104	396	122	-.511	.610	0.14	-.497	.619	0.01
Grammatical accuracy	486	126	503	120	531	131	-1.12	.259	0.13	-1.03	.301	0.35
Grammatical complexity	342	133	336	115	339	134	-.094	.925	0.04	-0.15	.978	0.02

Note. * $\alpha < .025$ (Bonferroni corrected)

FOREIGN LANGUAGE SPEECH LEARNING

Table 4

Partial Correlations Between the Pronunciation and Lexicogrammar Variables and Total Amount of L2 Input

Pronunciation variable	1 st Semester		2 nd Semester	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Segmental errors ^a	.30	.06	.10	.52
Word stress ^a	.46*	< .01	.33*	.04
Intonation ^a	.50*	< .01	.30	.06
Speech rate ^a	.36*	.02	.26	.11
Lexical appropriateness ^a	.36*	.02	-.01	.96
Lexical richness ^a	.37*	.02	-.08	.65
Grammatical accuracy ^a	.40*	.01	.04	.81
Grammatical complexity ^a	.37*	.02	.14	.38

Note. * $\alpha < .05$ ^aTheir proficiency scores at T1 and T2 were partialled out for the results of the first and second semesters, respectively.

FOREIGN LANGUAGE SPEECH LEARNING

Table 5
Partial Correlations Between the Pronunciation and Lexicogrammar Variables and Different Types of L2 Input

Pronunciation variable	1 st Semester						2 nd Semester					
	Content classes		Form classes		Extra use		Content classes		Form classes		Extra use	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Segmental errors ^a	.30	.06	.26	.10	.06	.71	.24	.13	-.05	.76	-.10	.52
Word stress ^a	.42*	<.01	.44*	<.01	.15	.33	.41*	<.01	.29	.07	-.05	.75
Intonation ^a	.45*	<.01	.42*	<.01	.24	.14	.42*	<.01	.14	.37	-.08	.61
Speech rate ^a	.35	.03	.36	.02	.14	.38	.38	.02	.11	.49	-.05	.72
Lexical appropriateness ^a	.40*	<.01	.36	.02	.01	.92	.05	.76	-.03	.83	-1.0	.54
Lexical richness ^a	.44*	<.01	.21	.20	.04	.78	-.03	.82	-.12	.46	-.01	.98
Grammatical accuracy ^a	.47*	<.01	.27	.09	-.05	.73	.02	.98	.07	.63	-.02	.86
Grammatical complexity ^a	.45*	<.01	.19	.24	.02	.89	.11	.49	.24	.13	-.11	.48

Note. * $\alpha < .012$ (Bonferroni corrected). ^aTheir proficiency scores at T1 and T2 were partialled out for the results of the first and second semesters, respectively.

FOREIGN LANGUAGE SPEECH LEARNING

APPENDIX: Details of the Individual Interview Protocol

<u>Name of class</u>	<u>How many hours per week?</u>					
<u>Classroom size (no. of students):</u>	<u>Instructor (e.g., natives, non-natives, Japanese)</u>					
<u>Class style (e.g., lectures vs. in-class discussion)</u>						
<u>How much did you use English during discussion/pair work in class?</u>						
I spoke very often 1	2	Neutral 3	4	5	I did not speak at all 6	7
<u>Who were your classmates that you frequently communicated with during class? Where were their nationalities from (natives, non-natives, Japanese)? In what situation, did you talk to them?</u>						
<u>How much did you think you understand what instructors said in class?</u>						
I understood him/her perfectly 1	2	Neutral 3	4	5	I did not understand at all 6	7