

Concurrent Validity of Sentence Repetition Test: The Role of Pause and Sentence Length

Midori SUNADA

Graduate School of Education Tokyo Gakugei University

Yuichi SUZUKI

University of Maryland

Abstract

In construction of Sentence Repetition Test (hereafter, SR), both sentence length and pause between the presentation of the target sentence and its repetition, have been found to play an important role in SR performance. The present study investigated how these two factors influence the task difficulty and the concurrent validity of SR with TOEIC scores. In order to achieve these aims, 79 Japanese high school students participated in this study and undertook the English SR under two conditions—with and without a pause. The SR sentences were varying lengths. The results showed that, while long sentences were more difficult to recall correctly than short ones, the pause made the short sentences more difficult and long sentences easier to repeat. Multiple regression analyses indicated that the performance under the most difficult condition, where the test-takers repeated long sentences without any pause, showed the highest concurrent validity with TOEIC. The implications of the present findings, along with some suggestions for further research, will be discussed to calibrate SR for second language test-takers.

1. Introduction

SR, which requires test-takers to listen to an English sentence and repeat it after a pause, has attracted attention from researchers as a measure of second language (L2) proficiency. SR performance has been reported to reflect the test-takers' proficiency (Erlam, 2006, Jessop, Suzuki, & Tomita, 2007). Three features are emphasized about SR. First, it requires both receptive and productive language skills. Second, when sentence repetition is attempted under time pressure, the achieved results are indicative of the learners' ability to respond to cue sentences presented at a natural speed. This exercise does not allow them to process the sentences slowly, as they need to respond quickly without being able to read the text. Third, sentence repetition has a history of being used as a test for L1 development, and it has been confirmed that this process is not parroting or simple rote repetition. The idea that sentence repetition is reconstructive has been supported by Ota (2009) and Endo (2010).

This technique, however, requires very careful calibration in designing an actual test (Jessop, Suzuki, & Tomita, 2007). For instance, sentence length and pause have been found to influence the performance of SR (McDade et al., 1982; Fraser et al., 1963,

Perkins, Brutton, & Angelis, 1986; Hendrickson, Aitken, McGhee, & Johnson, 2010). Few studies have examined the influence of task specifications of SR on the performance and concurrent validity, when administered alongside other standardized tests (e.g., TOEIC or TOEFL) (Sunada, 2014).

The current study aimed to clarify the role of parameters in SR in order to provide some valuable implications on the test development of SR.

2. Background

2.1 Overview of Sentence Repetition Test and Elicited Imitation Test

Sentence Repetition (SR) and Elicited Imitation (EI) have long been used as tools to measure the development of oral language skills in various contexts, including normal language development (Ervin-Tripp, 1964; Keller-Cohen, 1981; Menyuk, 1963), abnormal language development (Berry, 1976; Lahey, Launer, & Schiff-Myres, 1983; Menyuk, 1964) and oral language skills of second-language learners (Chaudron, Prior, & Kozok, 2005; Erlam, 2006; Jessop, Suzuki, & Tomita, 2007; Vinther, 2002). Although the meaning of the terms SR and EI may vary in different fields, we use the term SR to refer to both SR and EI as a measure of linguistic knowledge.

One criticism of SR is that it merely measures the short-term memory of test-takers, as it is inferred that repetition in SR does not require semantic and syntactic processing of the sentence. Many researchers, however, agreed that sentence repetition should include processes of decoding, interpretation, and reproduction of the stimulus sentences (Naiman, 1974; Ota, 2009; Endo, 2010).

Ota (2009) analyzed SR data using the participants' scores in various tests and tasks in order to confirm that the repetition of a sentence would be a cognitive task, involving semantic and syntactic processing. Japanese undergraduate and high school students participated in her study. One of her experiments included both SR and an oral composition (OC) test, which requires the participants to translate Japanese sentences into English orally. Ota conducted a qualitative analysis focusing on the errors the participants had made and phrases they had employed in both SR and OC. Similar errors and expressions were observed in the performance on these two tasks, suggesting that the process of SR is similar to that of producing a sentence. Ota made a further claim that SR performance reflects learners' grammar knowledge. This was supported by her experimental findings, where she asked participants to repeat English sentences followed by the task of translating the sentence into Japanese for each test item. The results showed that there was little correct repetition without correct translation into Japanese (i.e., comprehension of the sentence). This suggests that participants rarely succeed in repeating an English sentence without having comprehended it.

Endo (2010) replicated the work conducted by Ota (2009) by testing Japanese EFL junior and senior high school students whose English proficiency was lower than that the participants in Ota's study possessed. As his results indicated that few students

could repeat the sentences without comprehending them, they support Ota's (2009) finding that simple rote repetition rarely occurs and that SR is reconstructive in nature.

It is reported that SR reflects learners' overall language proficiency. Several researchers, such as Gallimore and Tharp (1981), Hamayan, Saegert, and Larudee (1977), and Munnich, Flynn, and Martohardjono (1994), claim that SR may be a good instrument to reflect learners' interlanguage representation. A study by Erlam (2006) examined the validity of SR as an implicit knowledge measure. It included the correlational analysis of the EI with a standardized English test (i.e., IELTS). The correlations between SR scores and each of the IELTS scores—for listening, speaking, reading, and writing—as well as their total IELTS scores, were calculated. A strong positive correlation was found between the SR scores and the IELTS total scores, indicating that overall L2 proficiency is reflected in their overall SR performance and possibly by learners' implicit grammar knowledge. Takeda (2010) collected the SR data and grammaticality judgment test (GJT) scores of Japanese undergraduate and graduate students and found a moderately high correlation between their scores in these tests. Overall, the accumulating evidence suggests that SR is a valid technique for assessing second language proficiency of a variety of participants.

2.2 Test Condition of Sentence Repetition

Although the validity of SR as a measure of L2 language knowledge has been confirmed in many previous studies, there are task specifications that need to be carefully defined in SR in order to avoid situations where test-takers can succeed in repeating without reconstructing the sentence. To avoid success by rote repetition, the length of the test item and the pause between the presentation of the target sentence and repetition should be carefully considered.

The pause between the presentation of the target sentence and its repetition is reported as crucial to avoid parroting in SR (McDade et al., 1982; Fraser et al., 1963). Delayed repetition, for example after a three-second pause, is said to produce more informative results about learners' language development than does immediate repetition (McDade et al., 1982). Under the delayed condition, learners have to understand the meaning of a given sentence, retain it, and rehearse it in their mind before repeating. In other words, due to the pause, learners cannot repeat a sentence by parroting (Fraser et al., 1963).

The length of the test items should also be well controlled. Some researchers agree that in SR, test-takers can repeat short sentences without understanding their meaning (Lee, 1970; Lust et al., 1987; Munnich & Martohardjono, 1994; Endo, 2011). As the sentence becomes longer, however, learners are required to divide the information into units, so that they can retain it in their working memory (Bley-Vroman & Chaudron, 1994).

Some researchers have empirically supported the idea that the length of the

sentences in SR is the prime determinant of the item difficulty (Perkins, Brutton, & Angelis, 1986; Hendrickson, Aitken, McGhee, & Johnson, 2010). Both Bailey et al. (1976) and Munnich et al. (1994), used sentences of 15 syllables when studying adult L2 learners. Graham et al. (2010) concluded that, in general, for adult testing, sentences should be 6–19 syllables in length. Hudgins and Cullinan (1978) found that sentences longer than 20 syllables were consistently difficult even for native speakers of English. Similar findings are drawn from the literature review on SR with both native speakers and second language learners (Vinther, 2002).

There have been several studies about the effects of the item length or the pause between the stimuli and repetition on the repetition performance. There have been few studies, however, about the effects of both the length and the pause on SR performance. The interaction effect of these two factors is thus examined in this work.

3. Design

3.1 Purpose

The purpose of this study is to investigate the influence of the item length of SR and the pause between the presentation of the target sentence and its repetition on SR performance. The results would contribute to testing the validity of SR as a measure of L2 knowledge.

In this study, the item length was controlled by the number of words in each sentence. In this test, sentences containing 5-7 words were considered short, while those of 9-11 words were longer sentences (see Material). In addition, two pause conditions were prescribed, labeled + pause and - pause. Under the + pause condition, the silence between the presentation of the target sentence and repetition was set at three seconds. Under the – pause condition, the participants were asked to repeat the sentence without a pause, immediately after they listened to the whole sentence.

The study examined whether these two factors, the sentence length and the pause, affect learners' repetition performance and the validity of SR as a measure of L2 knowledge. We chose TOEIC as a concurrent measure of L2 proficiency. TOEIC consists of only reading and listening sections and does not test productive skills, but comprehension ability is part of general L2 proficiency and related to productive skills (Sawaki, 2009). Besides that, each section of TOEIC has 100 questions, which test-takers should answer in a limited amount of time; 75 minutes for reading and 45minutes for listening. Test-takers need to process a large amount of text quickly. These led us to believe that TOEIC can be a good concurrent measure of the validity of SR. Both our SR and TOEIC measure fluency as well as accuracy.

3.2 Research Questions and Hypotheses

The study addresses two research questions as follows:

1. Do the sentence length and the pause between the stimulus and the repetition influence SR performance?
2. Does the concurrent validity of SR on TOEIC scores differ depending on the item length and the pause between the stimuli and the repetition?

The first research question asked how sentence length and pause influence the item difficulty of SR. Previous studies have shown that long sentences are more difficult to recall correctly than are short sentences (Perkins, Brutton, & Angelis, 1986; Hendrickson, Aitken, McGhee, & Johnson, 2010), and the current study also predicted that long sentences would be harder. According to the extant studies, the – pause condition with short sentences is likely to create the situation where students repeat the sentences like parroting, not reconstructing the sentence using their own linguistic knowledge (Lee, 1970; Lust et al., 1987; Munnich & Martohardjono, 1994; Endo, 2011). We hypothesized that the + pause condition would make the repetition more demanding because test-takers would need to retain the sentence in memory before they repeat it. In other words, the pause helps avoid parroting and makes it more conducive to reconstructive processing, which makes SR more demanding than the – pause condition. The predictions are summarized in Table 1.

Table 1

The Hypothesized Item Difficulty under the Different Conditions

		Pause	
		+ pause	– pause
Item Length	short	Easy	Very Easy
	long	Difficult	Medium

The second research question set out to examine the validity of SR as a measure of L2 proficiency under different conditions. The hypotheses on the concurrent validity of SR on TOEIC scores are presented in Table 2. The + pause condition with longer sentences would show the highest validity, as having a pause before repeating longer sentences would make it nearly impossible to repeat their full content correctly without reconstructing them. Thus, the performance under the + pause condition with longer sentences was hypothesized to reflect the test-takers' proficiency more clearly, in line with previous works (Lee, 1970; Lust et al., 1987; Munnich & Martohardjono, 1994; Endo, 2011). The – pause condition with short sentences was hypothesized to be the lowest in the concurrent validity on TOEIC scores because test-takers are sometimes able to repeat short sentences without a pause easily. This implies that the repetition under the – pause condition with short sentences does not necessarily require integrated linguistic knowledge. The concurrent validity of the other two conditions were

predicted to fall between the two conditions above. For the + pause condition with short sentences, test-takers were to retain the sentence they heard during the pause, which would make it harder to parrot the sentence. Similarly, for the – pause condition with long sentences, the long sentence would make it harder to parrot the sentence, even though no pause might not exclude the potential parroting.

Table 2

The Hypothesized Concurrent Validity of SR on TOEIC Scores under the Different Conditions

		Pause	
		+ pause	– pause
Item Length	short	Medium	Low
	long	High	Medium

3.3 Participants

The participants in this study were 79 senior high school students, 40 of whom were in the first year, and 39 were in the second year. Their average TOEIC score was 385.2 ($SD = 12.1$).

3.4 Material

Thirty-two English sentences were created for the study (see Appendix). Two lists were created (List A and List B), each with 16 sentences, four of which had 5, 7, 9, and 11 words, respectively. The sentences of five or seven words were simple sentences, while those containing nine or eleven words were complex. In reference to the previous studies (Bailey et al., 1976; Munnich et al., 1994; Graham et al., 2010), the number of syllables in a sentence ranged from 6 to 13 for short sentences, with a mean of 8.6. For long sentences, the number of syllables ranged from 11 to 17, with a mean of 13.4. We deliberately created complex sentences for the long test sentences. Long sentences with simple structure inevitably tax short-term memory (e.g., Nancy likes eating white chocolates, cookies, and colorful candies very much.), which makes SR a mere memory test. Therefore we used complex sentences for the long test sentences so as to measure proficiency.

Here are examples of short and long sentences.

Short: Bob drinks coffee after breakfast. (5 words, 8 syllables)

Tom read many books on summer vacation. (7 words, 10 syllables)

Long: While Kanako was cooking lunch, her mother watched TV. (9 words, 14 syllables)

I like my brother because he is kind to other people. (11 words, 15 syllables)

List A and List B were created such that the difficulty of sentences was kept as equal as possible. For both of these word lists, only those words that are familiar to Japanese learners of English were chosen. About 80% of the words were chosen from the 1000 word level of the JACET List of Basic Words (JACET stands for the Japan Association of College English Teachers) and the rest of the vocabulary was taken from COLUMBUS 21, which is one of the textbooks approved by the Ministry of Education. Other factors in the stimulus sentences that could potentially affect repetition performance were also controlled carefully. Only the present or past tense was used, with half of the sentences in present tense, and the other half in past tense. Likewise, third person singular *s* was only used in half of the stimulus sentences. The use of articles, *a* and *the*, was also controlled.

An American native speaker of English recorded her reading of the sentences. As previously noted, there were two conditions with respect to the pause between the presentation of the target sentence and the repetition: + pause and – pause conditions. For the + pause condition, the pause between the presentation of the target sentence and its repetition was set to three seconds, according to the report by McDade et al. (1982). For all the sentences, the time allowed for repeating their content was set at twice their duration. In other words, a sentence that took two seconds for the native speaker to read was allocated four seconds for repetition.

3.5 Procedure

In order to eliminate the practice effects due to the presentation order, the order of the two lists was counter-balanced. List A was administered to the group of first-year students under the – pause condition, and List B under the + pause condition. The second-year students took List A under the + pause condition and List B under the – pause condition. Before each test set, the participants were given four sentences as a practice. The participants took the test in class, and their repetition performances were recorded using computers.

3.6 Scoring and Analysis

Outliers in TOEIC score were detected based on the cutoff point of 3.0 in *z* score. Two participants were excluded from the analysis, due to their respective scores of (TOEIC = 740, *z* = 3.23) and (TOIEC = 935, *z* = 5.01). Thus, 77 of the original 79 participants provided the data that was used in the final analyses. The first author (a Japanese teacher of English who has a master degree of English education) and a native speaker of English counted the number of words the participants repeated correctly. The score was calculated by dividing the number of words repeated correctly in order by the

number of words in the original sentence (Endo, 2012). Twenty percent of the participants (i.e., 16 students) were randomly chosen, and their performance was scored by the two raters. The inter-rater reliability was high ($r = .93$). Other participants were divided into two groups, which were rated by the first author and the native speaker, respectively. Substitution of words was allowed as long as the meaning of the sentence was retained and grammatically correct. The omission of third person singular *s* and the articles *a* and *the* was scored as incorrect because it violates the grammaticality of the sentence. Some added words, both grammatical and ungrammatical, were not counted; they were ignored. For instance, “Tom read many books on *the* summer vacation” for “Tom read many books on summer vacation” got the full mark.

4. Results and Discussion

4.1 The Effect of Sentence Length and Pause on SR Performance

Mean scores and standard deviations for all the dependent variables for sentence length and pause are presented in Table 3. As can be seen, the score in the – pause condition is higher than the one in the + pause condition for the short sentences, whereas the score in the + pause condition is higher than that in the – pause condition for the long sentences. A two-way Analysis of Variance (ANOVA) with sentence length (short and long) and pause (+ pause and – pause) as within-subject factor was conducted. It yielded a main effect for the sentence length, $F(1, 76) = 1977.231$, $p < .001$, $partial \eta^2 = .963$, indicating that long sentences are more difficult to repeat than the short sentences are. The main effect of the pause was non-significant, $F(1, 76) = .110$, $p > .05$, $partial \eta^2 = .001$. Significant interaction effect was also revealed, $F(1, 76) = 12.361$, $p = .001$, $partial \eta^2 = .140$. The interaction indicates that the effect of the pause varies depending on the sentence length.

Table 3

The Effects of the Sentence Length and the Pause on SR scores

	+ pause		– pause	
	Mean	SD	Mean	SD
Short	.808	.122	.832	.102
Long	.417	.173	.381	.167

Two paired-sample t-tests were conducted to compare the SR score in the + pause and the – pause conditions separately for the short and the long sentences. In the short sentence condition, there was a marginally significant difference with a small effect size, $t(76) = 1.668$, $p = .099$, $d = .22$. In the long sentence condition, a similar magnitude of effect size was revealed between the two pause conditions, approaching a marginal statistical significance, $t(76) = -1.597$, $p = .115$, $d = .21$. These results suggest

that the three-second pause made SR more difficult when short sentences were used, while it made the test with long sentences easier. The lower scores in the + pause condition can be interpreted as indicating that the participants that were simply parroting without reconstructing the sentences found it more challenging to repeat the sentence in the + pause condition. Some of the short sentences might have been repeated without a pause by rote memory, whereas the three-second pause might have prevented repetition by rote memory. In contrast, it seems more difficult to repeat sentences without a pause in the long sentences. Without any preparation prior to repeating a sentence, processing and repeating its content may be more demanding. Participants might need to plan and repeat while they are repeating (i.e., online speech planning). It should be noted, however, that this interpretation must be considered with caution. The time allocated for the repetition was relatively short for the target population (i.e., high school students). As a result, in the – pause condition, the participants might have rushed to repeat the sentence, thus having less opportunity to reconstruct the sentence. In other words, a pause might have helped the participants pre-plan their repetition of the longer sentences.

4.2 Concurrent Validity of TOEIC scores with Sentence Length and Pause

In order to examine the concurrent validity of SR with the different conditions, multiple regression analysis was conducted on the total TOEIC score with the four SR scores as predictors. The omnibus test revealed that the model was significant, $F(4, 72) = 13.836, p < .001, R^2 = .435$, accounting for 43.5% of the variance in the TOEIC score.

Zero-order correlations between TOEIC and the SR scores are reported in Table 4. Each of the predictor variables correlated significantly with each other ($p < .01$). Regression coefficients in the multiple regression model are presented in Table 5. Tests for multicollinearity indicated that a very low level of multicollinearity was present for all the predictors; VIF was less than 10 and tolerance was above .02 (Field, 2009). The SR score based on long sentences in the – pause condition was the only significant predictor ($\beta = 0.532, p < .01$). None of the other predictors were statistically significant ($p > .1$).

Table 4

Zero-Order Correlations between TOEIC and SR scores

TOEIC	Short/ – pause	Long/ – pause	Short/ + pause	Long/ + pause
TOEIC	.523**	.639**	.290**	.338**
Short/– pause		.743**	.365**	.295**
Long/– pause			.399**	.353**
Short/+ pause				.783**
Long/+ pause				

** $p < .01$

Table 5

Predicting total TOEIC score

	Unstandardized Coefficients		Standardized Coefficients		t	p	Tolerance	VIF
	B	SE	β					
(Constant)	227.90	89.02			2.560	.013		
Short/ – pause	92.43	109.92	0.112		.841	.403	.442	2.262
Long/ – pause	267.38	68.19	0.532		3.921	.000	.427	2.344
Short/ + pause	-97.08	101.05	-0.141		-.961	.340	.365	2.741
Long/ + pause	110.39	69.43	0.227		1.590	.116	.384	2.607

We had stipulated that the + pause condition in which longer sentences were repeated would be the best predictor of the TOEIC score; however, our findings revealed that the long sentence with – pause was the best predictor of L2 proficiency. The standardized coefficient of long sentence with + pause was the second best predictor. In the short sentences, the – pause condition also turned out to be a better predictor than the + pause condition. These findings suggest that adding the pause did not make SR explain more variance of the TOEIC score. Why did giving no pause before repetition improve the concurrent validity of the proficiency score? In the – pause condition, time pressure was imposed on both comprehension and reconstruction processes of the sentences, and the test-takers had to comprehend and reconstruct the

sentences immediately. It may be possible to assume that this immediate processing required higher degree of proficiency than in the + pause condition where they had more time to comprehend and reconstruct the sentences. That is why the – pause condition showed more immediate relevance to the TOEIC scores. It is also plausible to assume, however, that the pause only helped the test-takers to comprehend the sentences but not to reconstruct them during the repetition phase. The test-takers might have used the extra time to comprehend the sentences, but the degree of their proficiency might not have been high enough to both comprehend *and* reproduce the sentences in real time. This interpretation is partially supported by the fact that the mean scores of repetition were not high and that test-takers' proficiency level indicated by the TOEIC scores was intermediate. In sum, in order for the test to have more validity, it may be better to set more demanding test conditions with this specific population (i.e. high school students).

5. Conclusion

This study aimed to address two research questions by investigating the influence of the item length of SR and the pause between the presentation of the target sentence and its repetition on SR performance and on the concurrence validity of SR as a measure of learners' proficiency. With respect to the first research question about the item difficulty under the different conditions of item length and the pause, in general, long sentences were more difficult to recall correctly than were short sentences. The pause between the presentation of the target sentence and its repetition, however, made the short sentences more difficult and long sentences easier to repeat. Contrary to our initial hypothesis that the most difficult condition was the + pause condition with longer sentences, it turned out to be the one where the participants were required to repeat longer sentences immediately after hearing them. However, because the effect sizes were small and only marginal significance was obtained, the generalizability of the pause effect must only be accepted with caution until the findings can be replicated.

The second question addressed the concurrent validity of SR with L2 proficiency measured with the TOEIC test. The performance under the most difficult condition, where test-takers repeated long sentences without any pause, showed the highest concurrent validity with TOEIC. This condition was the most demanding and reflected L2 proficiency measured with the TOEIC test because the time the test-takers had to understand the sentence was limited and online processing including pre-planning might have been required.

6. Suggestions for Further Studies

The current study opens several avenues for future research. First, in our experiments, the duration of the time allocated for repeating was twice as long as the length of each sentence. The recorded model utterance by the native speaker, however, was a little fast and the time for repeating it might have been short for the high school

students who participated in this study. Thus, the length of the pause can be manipulated in addition to the + pause and – pause condition. Furthermore, as the effects of pause may vary depending on item characteristics and test-takers' proficiency, further research should include a wide range of population, as well as test items, in order to better understand how the test condition of SR affects the validity of the task. Second, the present study focused on the item length of SR. However, the role of the pause seems to be influenced by the time allowed for repetition. Limiting the time allocated for repetition has been reported to make it possible to assess learners' automatic processing (e.g., Ellis, 2005b). Considering the item and the pause length and the time pressure to observe the automatic processing, the time for repetition should be controlled carefully not to be too short or too long. In future research, the effects of time pressure on the repetition can be investigated.

Acknowledgement

We would like to express our gratitude to the teachers and students who participated in this study. We are also grateful to Professor Tetsuo Baba at Tokyo Gakugei University for his comments on the earlier draft of this paper. Many thanks goes to a research assistant, Lani Alder, who scored the sentence repetition test.

References

- Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum International.
- Alderson, J. C. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Bailey, N., Eisenstein, M., & Madden, C. (1976). The development of wh-questions in adult second language learners. *On TESOL*, 76, 350-362.
- Berry, P. B. (1976). Elicited imitation of language: Some population characteristics. *Language and Speech*, 1, 350-362.
- Bley-Vroman, R., & Chaudron, C. (1994). Elicited imitation as a measure of second-language competence. *Research methodology in second language acquisition*, 7, 245-261.
- Chaudron, C., Prior, M., & Kozok, U. (2005). Elicited imitation as an oral proficiency measure. Paper presented at the 14th World Congress of Applied Linguistics, Madison, WI.
- Committee of Revising the JACET Basic Words (Ed.). (2003). *JACET list of 8000 basic words*. Tokyo: Japan Association of College English Teachers.
- Tomita, Y., Suzuki, W., & Jassop, L. (2009). Elicited imitation: Toward valid procedures to measure implicit second language grammatical knowledge. *TESOL QUARTERLY*, 43(2). 345-350.
- Ellis, R. (2005a). *Planning and task-performance in a second language*. Amsterdam:

John Benjamins.

- Ellis, R. (2005b). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, 27, 141-72.
- Endo, M. (2010). *The process of sentence repetition by Japanese EFL learners*. Unpublished graduation thesis. Tokyo Gakugei University, Tokyo.
- Endo, M. (2011). *The sentence repetition task and Japanese EFL learners*. Unpublished term thesis. Tokyo Gakugei University, Tokyo.
- Erlam, R. (2006). Elicited imitation as a measure of L2 Implicit knowledge: An empirical validation study. *Applied Linguistics*, 27(3), 464-491.
- Ervin-Tripp, S. M. (2010). Imitation and structural change in children's language. *New directions in the study of language*, 16, 163-189.
- Field, J. (2003). Promoting perception: Lexical segmentation in L2 listening. *ELT Journal*, 57, 325-334.
- Fodor, J. A., & Bever, T. G. (1965). The psychological reality of linguistic segments. *Journal of Verbal Learning and Verbal Behavior*, 4, 414-420.
- Fraser, C. U. B., & Brown, R. (1963). Control of grammar imitation, comprehension, and production. *Journal of Verbal Language and Verbal Behavior*, 2, 121-35. Reprinted in C. A. Ferguson & D. I. Slobin. (1973), *Studies of Child Language Development* (pp. 465-485) New York: Holt, Rinehart and Winston.
- Gallimore, R. & Tharp, R. G. (1981). The interpretation of elicited sentence imitation in a standardized context. *Language Learning*, 31(2), 369-392.
- Garrett, M., Bever, T., & Fodor, J. A. (1966). Active use of grammar in speech perception. *Perception and Psychophysics*, 1, 30-32. Cited in Kess, J. F. (1976). *Psycholinguistics: Introductory Perspectives*. New York: Academic Press.
- Graham, C. R., Lonsdale, D., Kennington, C., Johnson, A., & McGhee, J. (2008). Elicited imitation as an oral proficiency measure with ASR scoring. Retrieved August, 2010, from LREC Conferences Web site. (http://lrec-conf.org/proceedings/lrec2008/pdf/409_paper.pdf)
- Graham, C. R., McGhee, J., & Millard, B. (2010). The role of lexical choice in elicited imitation item difficulty. In M. T. Matthew, Y. Watanabe, & S. Lee (Eds.), *Selected Proceedings of the 2008 Second Language Research Forum: Exploring SLA Perspectives, Positions, and Practices*. Somerville, MA: Cacadilla Proceeding Project.
- Hamayan, E., & who. (1975). Elicited imitation in second language learners. *Working Papers on Bilingualism*, 6.
- Han, Y., & Ellis, R. (1998). Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research*, 2, 1-23.
- Hendrickson, R., Aitken, M., McGhee, J., & Aaron, J. (2010). What makes an item difficult? A syntactic, lexical, and morphological study of elicited imitation test items. *Selected Proceedings of the 2008 Second Language Research Forum*:

- Exploring SLA Perspectives, Positions, and Practices*, 48-56. Somerville, MA: Cascadilla Proceeding Project.
- Henning, G. (1983). Oral proficiency testing: Comparative validities of interview, imitation, and completion methods. *Language Learning*, 33(3), 315-332.
- Hudgins, J. C., & Cullinan, W. (1978). Effects of sentence structure on sentence elicited imitation responses. *Journal of Speech and Hearing Research*, 21, 809-819.
- Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.
- Hulstijn, J. (2002). Toward a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research*, 18, 193-223.
- Jessop, L., Suzuki, W., & Tomita, Y. (2007). Elicited Imitation in second language acquisition research. *The Canadian Modern Language Review*, 62(1), 215-238.
- Kellr-Cohen, D. (1981). Elicited imitation in lexical development: Evidence from a study of temporal reference. *Journal of Psycholinguistic Research*, 10(3), 273-288.
- Lahey, M., Launer, P., & Schiff-Myers, N. (1983). Prediction of production; Elicited imitation and spontaneous speech productions of language disordered children. *Applied Linguistics*, 14, 317-343.
- Lee, L. L. (1970). A screening test for syntax development. *Journal of Speech and Hearing Disorder*, 35(2), 102-12.
- Lust, B., Chien, Y., & Flynn, S. (1987). What children know: Methods for the study of first language acquisition. In B. Lust, *Studies in the Acquisition of Anaphora*, 2. Dordrecht: D. Reidel. 271-356.
- McDade, H. L., Simpson, M. A., & Lamb, D. E. (1982). The use of elicited imitation as a measure of expressive grammar: A question of validity. *Journal of Speech and Hearing disorder*, 47(1), 19-24.
- Menyuk, P. (1963). A preliminary evaluation of grammatical capacity in children. *Journal of Verbal Learning and Verbal Behavior*, 2, 429-439.
- Menyuk, P. (1964). Comparison of grammar of children with functionally deviant and normal speech. *Journal of Speech and Hearing Research*, 7, 109-121.
- Munnich, E., Flynn, S., & Martohardjono, G. (1994). Elicited imitation and grammaticality judgment tasks: What they measure and how they relate to each other. In E. Tarone, S. M. Grass, & A. Cohen (Eds.), *Research methodology in second language acquisition* (pp. 227-234). Hillsdale, NJ: Lawrence Erlbaum.
- Naiman, N. (1974). The use of elicited imitation in second language acquisition research. *Working Papers on Bilingualism*, 2, 1-37.
- Ota, E. (2009). *The role of sentence repetition in foreign language learning: Sentence repetition processes by Japanese EFL learners*. Unpublished doctoral dissertation, The United Graduate School of Education, Tokyo Gakugei

University, Tokyo.

- Perkins, K., Brutton, S. R., & Angelis, P. J. (1986). Derivational complexity and item difficulty in a sentence repetition task. *Language Learning*, 36, 125-141.
- Pienemann, M. (1998). *Language processing and second language development: Processability theory*. Amsterdam: John Benjamins.
- Sawaki, Y., Stricker, L. J., & Oranje, A. H. (2009). Factor structure of the TOEFL Internet-based test. *Language Testing*, 26(1), 5-30.
- Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson & D. I. Slobin (Eds.), *Studies of Child Language* (pp. 175-208). New York: Hilt, Rinehart and Winston.
- Slobin, D. I., & Welsh, C. A. (1968). Elicited imitation as a research tool in developmental psycholinguistics. Retrieved October 16, 2009, from Education Resources Information Center (ERIC) database.
- Sunada, M. (2014). Sentence repetition as a measure of learners' English proficiency. *Journal of Educational Research*. 29, 41-59. The United Graduate School of Education, Tokyo Gakugei University, Tokyo.
- Takeda, N. (2010). *Toward more practical measurement of implicit grammatical knowledge: A correlational study of elicited imitation test and time-oriented test*. Unpublished master's dissertation. Tokyo Gakugei University, Tokyo.
- Yamaoka, T. (2006). On the importance of imitation and repetition in foreign language learning. *Annual Review of English Language Education in Japan (ARELE)*, 17, 1-10.
- Vinther, T. (2002). Elicited imitation: A brief review. *International Journal of Applied Linguistics*, 12, 54-73.
- Wong, B. E. (2012). Elicited imitation as a measure of L2 English learners' interlanguage representation of relative clause. *Electronic Journal of Foreign Language Teaching*, 9(1), 91-107.

Appendix

The Sentences for SR

Item Number	Test Set	Sentence	Number of Words
1	1	My farther speaks French well.	5
2	1	I often play the guitar.	5
3	1	John played baseball after school.	5
4	1	Kyoko went to a restaurant.	5
5	1	I have a test in the morning.	7
6	1	Mary practices the piano in the morning.	7
7	1	Naoko sang her favorite songs at home.	7
8	1	Tom read many books on summer vacation.	7
9	1	We bring our lunches when we have a meeting.	9
10	1	Emi always does her homework before she eats dinner.	9
11	1	While Kanako was cooking lunch, her mother watched TV.	9
12	1	Mike joined the basketball club when he entered school.	9
13	1	Jack plays the guitar when he has free time on weekends.	11
14	1	Some children wash their hands after they come home from school.	11
15	1	Taro went to bed at eight because he was very tired.	11
16	1	Nancy did not buy a camera because she had little money.	11
17	2	Bob drinks coffee after breakfast.	5
18	2	I have a class today.	5
19	2	I visited the house yesterday.	5
20	2	We studied English very hard.	5
21	2	I clean my small room every day.	7
22	2	Misaki teaches a science class at school.	7
23	2	I saw many dogs in the park.	7
24	2	My brother opened the door for me.	7
25	2	I want an expensive computer when I get older.	9
26	2	Kenji always takes a shower after he eats breakfast.	9
27	2	After my friends passed the test, I cooked dinner.	9
28	2	When I called my friend, he was watching TV.	9
29	2	Satoshi buys fruit at the shop when he is very hungry.	11
30	2	I like my brother because he is kind to other people.	11
31	2	My father gave me a birthday present when I came home.	11
32	2	You ate my cake while I was talking to my mother.	11