Effect of Time Pressure on Grammaticality Judgment Tests with L1 Translation

Miki TOKUNAGA
Fukuoka University Language Education & Research Center
E-mail: tokunagamiki@fukuoka-u.ac.jp

Annual Review
of
English Learning and Teaching
No.21

The JACET Kyushu-Okinawa Chapter
November 30, 2016

－Article－

Effect of Time Pressure on Grammaticality Judgment Tests with L1 Translation

Miki TOKUNAGA
Fukuoka University Language Education & Research Center
E-mail: tokunagamiki@fukuoka-u.ac.jp

Abstract

Timed and untimed grammaticality judgment tests (GJTs) with L1 (Japanese) translations were given to Japanese university EFL learners ($N = 219$) to examine the effect of time pressure on item difficulties. The results of *t*-tests and factor analysis indicated that, for this group of participants, time pressure did not significantly affect the results. On the other hand, the difference in difficulties between grammatical items and ungrammatical items were larger and statistically significant, suggesting that grammatical and ungrammatical items in GJTs may measure different types of knowledge or ability of learners. Also, untimed grammatical items did not correlate with other item types and loaded heavily on the second factor. This could indicate that untimed grammatical items may be measuring something different, or they are simply much easier than other item types, causing them to show up as though they are measuring something different.

Keywords: GJT, time pressure, Rasch analysis

Introduction

Grammaticality judgment tests (GJTs) have often been used in SLA research to measure learners' implicit and explicit knowledge of the target language both abroad and in Japan (e.g., Ellis, 2009; Green & Hecht, 1992; Roeher, 2008; Sakai, 2008; Shimada, 2010). GJTs consist of a series of either grammatical or erroneous sentences. Test takers are asked to determine the grammaticality of each sentence, and in many cases, asked to do one or more of the followings: correct the error, explain the error using appropriate metalanguage, and indicate whether the decision was based on feeling or explicit knowledge.

GJTs are conducted with or without time pressure, and timed GJTs are thought to measure constructs related to implicit knowledge of the target language, while untimed GJTs are often presumed to measure constructs of explicit knowledge. Even though clear distinctions of constructs measured by timed and untimed GJTs are still being disputed, many previous studies agree that the tests measure different factors (e.g., Ellis, 2005; Godfroid, et al., 2015; Zhang, 2015). However, using the modified version of GJTs from Ellis (2005) with Japanese university EFL learners, Shimada (2010) found that the difference in mean scores was larger between grammatical and ungrammatical items than between timed and untimed items. He also found that timed and untimed items correlated more strongly than grammatical and ungrammatical items, suggesting that grammaticality of items affected the

scores more than time pressure. Gutiérrez (2013) presented a similar conclusion that, even though both time pressure and grammaticality of items significantly affected participants' performance, grammaticality of items had a stronger effect than time pressure.

Current study

This paper reports the results of pilot tests conducted as part of a larger study investigating the English grammar knowledge and performance of Japanese university EFL learners. The main project will involve a grammar rule test, written and oral sentence translation tests, and written and oral picture description tests, as well as timed and untimed GJTs. GJTs in this project differ from those in previous studies in two aspects: adding L1 (Japanese) translations for target sentences, and using the Rasch model for analysis. The purpose of this current study was to pilot the method, items, and types of analysis, to find out whether time pressure in these GJTs and analysis would show similar results with previous studies, thus indicating that timed and untimed GJTs possibly measure different factors of learners' L2 knowledge or ability.

Participants

The participants of this study were 260 students at a private Japanese university. They were either first or second year students taking required English classes, and none were English majors. All participants' L1 was Japanese. Data from 219 participants who signed the consent form and took both the timed and untimed GJTs were used for analysis in this paper. The two tests were administered in two separate class periods more than two weeks apart. They were conducted as part of the coursework with explanations of answers following the untimed GJT to review the grammar points covered in the tests.

Items

Four different forms of GJTs were designed, with each form consisting of 37 items from a 68 item pool covering 20 grammar structures listed in Table 1. The 20 grammar structures were chosen based on Ellis (2005) and Shimada (2010), plus three added structures: *go* + *ing*, prepositions and "subject". *Go* + *ing* refers to errors of adding *to* before gerunds. Japanese students often add *to* after *go* regardless of what follows *go*, creating such an erroneous sentence as "*I went to camping last weekend." An example of the "subject" structure item is "*Today has three classes." The accompanying Japanese translation is「今日は授業が３つある。」, which means "I have three classes today." Because subjects can be omitted in Japanese, it can be difficult for Japanese EFL learners to add an appropriate subject when making such sentences in English.

All sentences were written for this study, and the vocabulary used was limited to Level 2 on JACET 8000 (Ishikawa et al., 2003) to keep the effect of participants' vocabulary knowledge to a minimum. The major difference between the GJTs used in this study and previous studies was

that the tests in this study included Japanese (L1) translations for all items.

Ellis (2004) states that GJTs potentially involve three processing operations: understanding the meaning of a sentence (semantic), deciding whether something is formally incorrect in the sentence (noticing), and considering what is incorrect and why (reflecting). However, the first step involves more than the test takers' grammatical knowledge, and for EFL learners of beginner to intermediate levels of English, the first step could prevent them from moving to step two, where they are to judge the grammaticality of the sentence. By adding a Japanese translation to each sentence, this test attempted to minimize the effect of learners' vocabulary knowledge and reading ability and measure their understanding of target grammar structures. An example item from GJTs used in this study is shown below:

* Does your mother a teacher? あなたのお母さんは先生ですか。

Table 1. *Grammatical Structures Included in the GJTs*

Ellis (2005)	Shimada (2010)	Current study
Verb complement	Verb complement	Verb complement
Regular past	Regular past	Regular past
Questions tags	Questions tags	Question tags
Yes / no questions	Yes / no questions	Yes / no questions
Modal verbs	Modal verbs	Modal verbs
Unreal conditionals	Unreal conditionals	Unreal conditions
Since and *for*	*Since* and *for*	*Since* and *for*
Indefinite article	Indefinite article	Indefinite article
Possessive –s	Possessive –s	Possessive –s
Plural –s	Plural –s	Plural –s
Third person –s	Third person –s	Third person –s
Relative clauses	Relative clauses	Relative clause
Embedded questions	Embedded questions	Embedded questions
Comparatives	Comparatives	Comparatives
Adverb placement	Adverb placement	Adverb placement
Ergative verbs	Ergative verbs	Conjunctions
Dative alternation	Dative alternation	Subject
	Reported speech	Reported speech
	Progressive	Prepositions
	Irregular past	*Go + ing*

There were two ungrammatical items for each of the 20 grammar structures (40 items), and two grammatical items for 11 of the structures (22 items). The reason for only having 11 grammatically correct structures was to make each test relatively short, so the participants could stay focused throughout the tests. Those 62 items were divided into four short test forms consisting of 31 items, with each of the 62 item appearing in two of the four forms, allowing the data to be merged later using those shared items. Despite the four forms having shared items, forms 1 and 2 did not have shared items, nor forms 3 and 4. Participants who

took form 1 in the timed GJT was given form 2 in untimed GJT, and participants who took form 3 in the timed GJT were given from 4 in untimed GJT. In addition, there were six ungrammatical items covering six of the grammar structures used as common items, appearing in all four forms of the tests. Thus, there were 37 total items in each test form. Although the participants encountered the 6 common items in both GJTs, the answer was not given after the timed GJT and there were at least two weeks between the two tests. The data from the 4 forms were merged using the overlapping and common items.

Timed GJT

The timed GJT was conducted with items in a PowerPoint presentation projected onto a screen. Each sentence with a Japanese translation was shown for 10 seconds, followed by 5 seconds of blank page to allow participants to look down and mark their answer on the answer sheet. Participants were asked to judge the grammaticality of each English sentence and mark either grammatical, ungrammatical, or not enough time to judge. The 10 second time limit was arbitrarily set by the author after consulting previous research (Gutiérrez, 2013; Kusanagi, 2012; Loewen, 2009; Shimada, 2010; Zhang, 2015). For example, Loewen (2009) pretested his 68 items to 20 L1 English speakers, calculated median time took for each item, and added 20% to give extra time for L2 learners. However, 140 L2 participants could not answer an average of 12 items out of the 68 in time, indicating that the time given was not enough for those missed items. There seems to be no clear way of determining the right amount of time for timed GJTs, and to set it properly, each time limit needs to be catered to each sentence length and each participant's ability, which seemed beyond the scope of this pilot test. Ten seconds seemed long enough for participants to finish reading two (English and Japanese) short sentences, but short enough to give a sense of pressure.

Untimed GJT

The untimed GJT was conducted as a pencil and paper test, in which participants were asked to judge the grammaticality of each sentence. As the second part of the test, for the sentences that they marked as ungrammatical, they were asked to correct the errors. However, the results of this correction part are not reported in this paper. The test was started 30 minutes before the end of the class period, and the participants were required to stay for 20 minutes. After the initial 20 minutes, participants who had completed all the items were allowed to leave the classroom. In all groups, only a few students had to stay beyond the first 20 minutes, and no one needed more than 30 minutes.

Results

Rasch analysis using the Winsteps® software package (Linacre, 2016b) was conducted on the data from the tests. Raw scores give an ordinal measurement, which is rank-ordered, and the distance between the ranks is ignored. For example, the difference in abilities

between someone who answered 10 out of 100 questions correctly and another person who correctly answered 30 is not necessarily equal to the difference between two other people who correctly answered 80 and 100 questions, even though the difference in number of questions are both 20. This is because difficulties of all questions are not equal. Rasch analysis produces measures of item difficulty and person ability on a common equal interval scale measured in log-odds units (logits) with mean item difficulty set as 0 logits by convention. This means that the distance between each interval on the scale is equal.

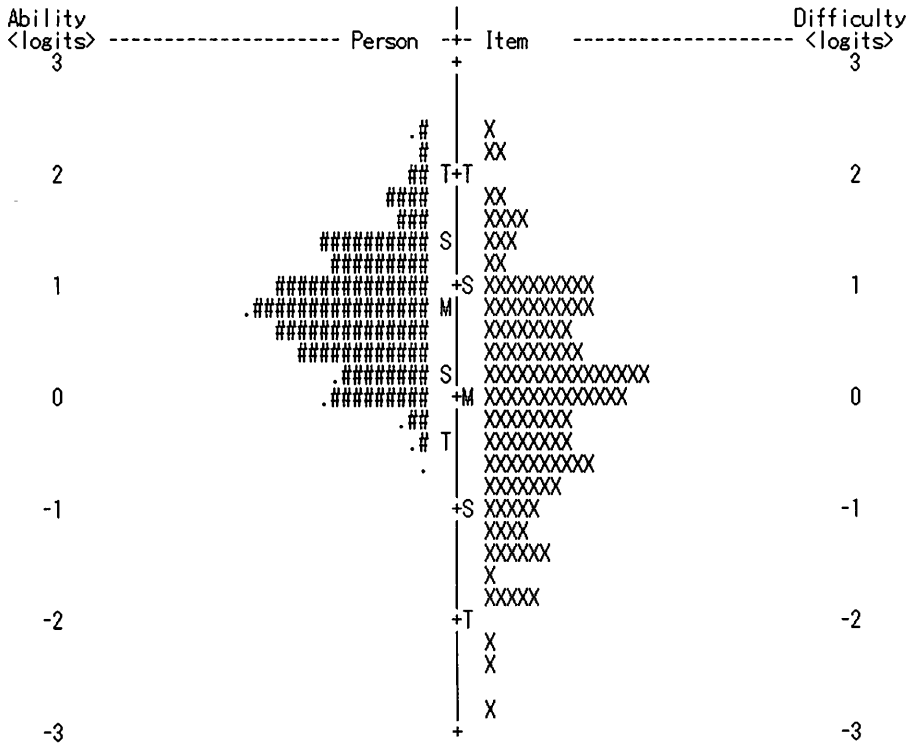

Figure 1. Variable map of the combined GJT showing abilities of 219 participants on the left and difficulties of 136 item on the right. Each "#" represents two persons, each "." equals one person, and each "X" is one item.

Data from timed and untimed GJTs were merged using common and shared items. Thus, 68 items from each of the two tests were treated as different items in one merged test of 136 items. Figure 1 shows the Winsteps variable map of the combined GJT with persons ranked by ability on the left and items by difficulty on the right. Each "X" represents one item, while "." and "#" represent one and two participants respectively. A position higher on the scale represents greater ability for a person or greater difficulty for an item. When person ability precisely matches item difficulty, the person has a 50% expectation of success.

The M on each side of the axis shows the mean of person ability and item difficulty, and it is shown that participants' mean ability was higher than the mean item difficulty of items on the GJTs. It is also clear on the map that there were no participants matching the difficulties of the easiest items. As this test was intended to measure understanding of grammar structures by low to intermediate-proficiency learners, the scarcity of more difficult items and the existence of items that were too easy for this group does not indicate problems of the instrument. What this means is that structures of items low on the variable map were understood even by participants with low ability.

In the timed GJT, 35 responses indicated that the participants did not have enough time to make a judgment on the grammaticality of the items. Because of the small number (0.4% of all timed responses), these responses were counted as incorrect for purposes of comparability with previous research (Gutierrez, 2013).

Table 2 shows the summary statistics of the combined GJT. As mentioned above, mean item difficulty is conventionally set to 0 logits in Rasch analysis, resulting in mean item difficulty to be 0.00 logits. In Rasch analysis, person and item reliabilities do not report on the quality of the data. Rather, they show the reproducibility of the results. Thus, a high reliability coefficient means that persons (or items) estimated to have high measures actually do have higher measures than persons (or items) estimated with low measures (Linacre, 2016a). Person separation and item separation show how many classes the persons and the items can be separated into by the test.

Table 2. *Summary Statistics for the Combined Analysis*

	N	*M* (logit)	*SE*	*SD*	Separation	Rasch Reliability
Items	136	0.00	0.23	0.95	3.95	.94
Persons	219	0.79	0.28	0.52	1.83	.77

Item reliability (.94) is higher than person reliability (.77) as there were more participants to measure the difficulty of the items than items to measure abilities of participants. Person separation (< 2.0) shows that the test is not sensitive enough to distinguish between high and low performers. On the other hand, item separation (> 3) shows that there are enough participants to confirm the item difficulty hierarchy. As the main purpose of this research project is to measure the difficulties of different grammar structures and not to rank the participants' L2 levels, the low person reliability and separation do not significantly affect the quality of the research.

Table 3 displays mean item difficulties for timed, untimed, grammatical and ungrammatical items from the combined analysis. It shows that untimed items (M = -0.10 logits) were easier than timed items (M = 0.10 logits), and grammatical items (M = -0.59 logits) were much easier than ungrammatical items (M = 0.28 logits), with ungrammatical item being the most difficult of the four item types and grammatical items the easiest.

Table 3. *Item Difficulties for Item Types*

Items	Item difficulties		
	n	*M* (logit)	*SD*
Timed	68	0.10	0.86
Untimed	68	-0.10	1.09
Grammatical	44	-0.59	0.91
Ungrammatical	92	0.28	0.89

Table 4 lists items of the 15 highest and 15 lowest difficulties out of 136 items (68 items in timed and untimed tests). It is easily seen that all 15 most difficult items were ungrammatical sentences, and most of the easiest items were grammatical sentences, indicating that judging grammatical items as correct was easier than judging ungrammatical items as incorrect. Also, there are more timed items in the 15 most difficult list, and more untimed items in the easiest list, suggesting that time pressure had made the items more difficult. However, out of 68 different sentences, both timed and untimed versions of the same five sentences (*You should wear hat today; * I saw very funny movie last night; * Do you know when is her birthday?; * I would buy it if it is not so expensive; and * I know who are you.) appear twice in the list of 15 most difficult items, and timed and untimed versions of the same three sentences (Bob decided to learn French; Are you hungry?; and * Takuya walk home last night.) appear in the list of easiest items. This could mean that time pressure did not have a big effect on the difficulties of these items.

Table 4. *Items of Highest and Lowest Difficulties*

Highest difficulties			Lowest difficulties		
Logit	Time	Item	Logit	Time	Item
2.33	T	*You should wear hat today.	-1.32	U	I was tired, so I didn't do my homework.
2.21	U	*I saw very funny movie last night.	-1.33	U	It isn't raining, is it?
2.15	U	*You should wear hat today.	-1.36	U	Bob decided to learn French.
1.8	U	*Today has three classes.	-1.41	T	Bob decided to learn French.
1.79	U	*Do you know when is her birthday?	-1.43	U	She ate dinner, didn't she?
1.64	T	*I saw very funny movie last night.	-1.44	U	Is your mother a teacher?
1.55	U	*I would buy it if it is not so expensive.	-1.69	U	I can't buy that because I have no money.
1.52	T	*Do you know when is her birthday?	-1.74	U	We hope to see you at the party.
1.51	U	*I went to camping last weekend.	-1.81	U	I'll see you on Friday.
1.49	T	*I would buy it if it is not so expensive.	-1.81	U	Are you hungry?
1.39	T	*Does your mother a teacher?	-1.82	T	*Takuya walk home last night.
1.32	U	*I know who are you.	-1.87	U	The man who wrote the book is my father
1.24	T	*I know who are you.	-2.11	U	*Takuya walk home last night.
1.18	U	*What is the teacher name?	-2.44	T	Are you hungry?
1.05	T	*Jane likes to go to swimming on weekends.	-2.75	U	Jane was born in 1997.

Note. T = timed; U = untimed; * = ungrammatical sentence

To compare means of timed and untimed items and grammatical and ungrammatical items, two-sample *t*-tests were conducted. The reason for using two-sample *t*-tests was because participants took different forms of GJTs for timed and untimed tests. A particular item's timed and untimed version was taken by different participants, except for common items. A preliminary test for the equality of variances indicated that the variance of the timed and untimed items were found to be different ($F = .66$, $p = .03$). Therefore, a two-sample *t*-test with unequal variance was performed for this group. On the other hand, the variance of the grammatical and ungrammatical groups found to have an equal variance ($F = 1.51$, $p = .40$), so a two-sample *t*-tests assuming equal variance was performed for this group.

The mean difficulty for timed items ($M = 0.10$, $SD = 0.86$) was 0.20 logits higher than the mean difficulty for untimed items ($M = -0.10$, $SD = 1.09$). However, the difference was not statistically significant ($p = .25$), with an effect size of .10, which is considered small (Cohen, 1988). On the other hand, the mean difficulty for grammatical items ($M = -0.59$, $SD = 0.91$) was 0.87 logits lower than ungrammatical items ($M = 0.28$, $SD = 0.89$), and the difference was statistically significant with medium effect size ($p < .001$, $r = .44$).

Table 5. *Results of Two-Sample t-tests*

	N	*M*	*SD*	Two-sample *t*-test			
				t	*df*	*p*	*r*
Timed	68	0.10	0.86	1.15	127	.25	.10
Untimed	68	-0.10	1.01				
Grammatical	44	-0.59	0.91	-5.32	134	.00	.44
Ungrammatical	92	0.28	0.89				

To examine the effect of time pressure and item grammaticality a little more closely, mean item difficulties were divided into upper and lower ability groups of participants by person ability measures (Table 6). Participants with person ability of 0.73 logits and below were placed in the lower ability group ($n = 109$), and those with person ability of 0.77 logits and above were placed in the upper group ($n = 110$). The cut-off point was decided so the two groups would consist of similar number of participants. The difference in timed and untimed items for lower ability group is 0.14 logits, while it is 0.26 logits for the upper group, suggesting that the extra time in untimed test did not help the lower group as much as it did the upper group. On the other hand, the difference in mean difficulties for grammatical and ungrammatical items is 1.06 logits for the lower group and 0.38 logits for the upper group, indicating again that grammaticality of the item had a bigger effect on difficulty of items.

Table 6. *Mean Difficulties for Ability Groups*

	n	Mean difficulties (logit)				Difference in mean difficulties (logit)	
		T	U	Gr	UGr	T vs. U	Gr vs. UGr
Lower	109	0.07	-0.07	-0.72	0.34	0.14	1.06
Upper	110	0.10	-0.16	-0.19	0.19	0.26	0.38
Whole	219	0.10	-0.10	-0.59	0.28	0.19	0.87

Note. T = timed; U = untimed; Gr = grammatical; UGr = ungrammatical

To examine the correlations between item types, average person abilities of participants for different item types were used. Table 7 presents the correlation matrix between person abilities from four types of items: timed grammatical items (Timed Gr), timed ungrammatical items (Timed UGr), untimed grammatical items (Untimed Gr), and untimed ungrammatical items (Untimed UGr). Correlations were statistically significant between all combinations except for three combinations including untimed grammatical items: Untimed Gr and Timed Gr, Untimed Gr and Timed UGr, and Untimed Gr and Untimed UGr. This suggests that untimed grammatical items might be measuring a different factor to the rest of the item types.

Table 7. *Correlation Matrix of Person Abilities between Item Types*

	Timed Gr	Timed UGr	Untimed Gr	Untimed UGr	All
Timed Gr	—				
Timed UGr	.39**	—			
Untimed Gr	.09	.05	—		
Untimed UGr	.19*	.56**	.06	—	
All	.50**	.84**	.24**	.84**	—

Note. $^{*}p < .05$ (2-tailed); $^{**}p < .001$ (2-tailed); Gr = grammatical; UGr = ungrammatical

Looking at correlations with the person abilities against all items, abilities against timed and untimed ungrammatical items have bigger effect sizes ($r = .84$ for both timed and untimed) than abilities against grammatical items ($r = .50$ for timed and $r = .24$ for untimed). This may suggest that the ability to judge ungrammatical items can affect the total ability more than the ability to judge grammatical items, both in timed and untimed conditions.

Finally, following Gutiérrez (2013), a principal components factor analysis with direct oblimin rotation was conducted with SPSS 23. Unlike the findings of Gutiérrez (2013), only one factor was found based on eigenvalues. By extracting 2 factors, different types of items loaded as shown in Table 8. Both timed and untimed ungrammatical items loaded heavily on the first factor, emphasizing the earlier observation that ungrammatical items have a bigger effect on the performance on the GJT than grammatical items. Timed grammatical items also loaded on the first factor, while untimed grammatical items alone

loaded heavily on the second factor. Along with its insignificant correlations with other item types on Table 7, this result could again suggest that untimed grammatical items may be measuring something different from other types of items.

Table 8. *Loadings for Principal Component Factor Analysis*

Items	Component 1	Component 2
Timed Gr	.59	.21
Timed UGr	.89	-.07
Untimed Gr	-.01	.98
Untimed UGr	.80	-.12

Note. Rotation method = Direct oblimin; Gr = grammatical; UGr = ungrammatical.

Discussion and Conclusion

This short paper examined whether or not timed and untimed GJTs could be used to measure different types of L2 knowledge or ability. In this study, both timed and untimed GJT sentences had L1 (Japanese) translations to minimize the effect of participants' vocabulary knowledge on their judgment on grammaticality.

The Winsteps variable map (Figure 1) showed that the mean difficulty of the GJT items in this study were lower than the mean ability of this study's participants. Having Japanese translations may have made the tasks easier, by reducing the effect of participants' vocabulary knowledge and reading ability on their grammaticality judgment. Further tests need to be conducted to investigate the effect of L1 translations: whether it makes all items easier or it only helps certain grammar structures. The item reliability of the combined analysis was high (Table 2), and the item separation of 3.95 indicated that the test had enough participants to measure the item difficulty hierarchy.

Item difficulties for item types (Table 3) showed that the easiest items were grammatical items, followed by untimed items and timed items, with ungrammatical items being the most difficult. In fact, there were more ungrammatical items in the 15 most difficult items and more grammatical items in the 15 easiest items (Table 4), supporting the above findings that judging ungrammatical items as incorrect was more difficult than judging correct sentences as correct. This was confirmed by two-sample *t*-tests (Table 5), which showed that the difference between the means of grammatical and ungrammatical items was significant with a larger effect size than that of the comparison between timed and untimed items. The correlation matrix (Table 7) showed that untimed grammatical items did not significantly correlate with other item types, indicating that this type of items could be measuring something different than what other item types were measuring. Supporting this finding, exploratory factor analysis showed that untimed grammatical items could be a different component (Table 8).

This pilot study showed that, even though both time pressure and grammaticality

affected the difficulty of items, the effect of grammaticality was more significant than that of time pressure, indicating that grammatical and ungrammatical items on GJTs may measure different factors of learner's L2. Among grammatical items, untimed grammatical items stood out to be different from other items in correlation and factor analysis. Further analysis is required to find out whether different factors are measuring different types of knowledge, as previous studies claimed, or they are simply different levels of difficulties. GJTs for the main study will be revised according to results discussed here, along with results of other pilot tests.

Acknowledgment

This work was supported by Grant-in-Aid for Scientific Research (JSPS KAKENHI Grant Number 26370754).

References

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning, 54,* 227-275.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition, 27*(2), 141-172.

Ellis, R. (2009). Measuring implicit and explicit knowledge of a second language. In R. Ellis, S. Loewen, C. Elder, J. Erlam, J. Philp., & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 31-64). Bristol: Multilingual Matters.

Godfroid, A., Loewen, S., Jung, S., Park, J., Gass, S., & Ellis, R. (2015). Timed and untimed grammaticality judgments measure distinct types of knowledge: Evidence from eye-movement patterns. *Studies in Second Language Acquisition, 31,* 269-297.

Green, P., & Hecht, K. (1992). Implicit and explicit grammar: An empirical study. *Applied Linguistics, 13,* 168-184.

Gutiérrez, X. (2013). The construct validity of grammaticality judgment tests as measurers of implicit and explicit knowledge. *Studies in Second Language Acquisition, 35,* 423-449.

Ishikawa, S., Uemura, T., Kaneda, M., Shimizu, S., Sugimori, N., & Tono, Y. (2003). *JACET8000: JACET list of 8000 basic words.* Tokyo: JACET.

Kusanagi, K. (2012). Jikan seigen wo mochiita bunnpousei hanndan kadai - kisoteki kentou to jikan seigen no settei houhou ni tuite. [Grammaticality judgment task with time pressure; Survey of the test format and time limit setting method]. *LET Kansai Chapter Methodology SIG Journal, 2012,* 46-67.

Linacre, J. M. (2016a). *A user's guide to Winsteps Ministep Rasch-model computer*

programs 3.92.0. Retrieved from http://www.winsteps.com.

Linacre, J. M. (2016b). Winsteps (Version 3.92.1) [Computer software]. Retrieved from http://www.winsteps.com.

Loewen, S. (2009). Grammaticality judgment tests and the measurement of implicit and explicit L2 knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 94-112). Bristol: Multilingual Matters.

Roehr, K. (2008). Metalinguistic knowledge and language ability in university-level L2 learners. *Applied Linguistics, 29,* 173-199.

Sakai, H. (2008). Implicit and explicit grammatical knowledge of L2 English: Identification, correction, and provision of rules. *Annual Review of English Language Education in Japan, 19,* 91-100.

Shimada, K. (2010). Bunpousei hanndan tesuto ni okeru monndai teiji jikannseigen no umu to meijiteki annjiteki chishiki. [Timed/timed grammaticality judgment test and implicit /explicit knowledge]. *St. Andrew's University English Review, 24,* 41-53.

Zhang, R. (2015). Measuring university-level L2 learners' implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition, 37,* 457-486.