論 文

Using the Spacing Effect with Multiple-choice Items on a Listening Test

Paul Horness

Listening Comprehension

Rost (2002) defined listening as the "mental process of constructing meaning from spoken input" (p. 279). This complex process involves a continuum of active processes, which the listener controls, and passive processes, which the listener does not control. The active processes can be divided into two levels: recognizing the auditory input and understanding its meaning. This process is often referred to as bottom-up and top-down listening. Bottom-up listening means using the information about sounds, word meanings, and discourse markers to attain understanding of what is heard one step at a time. Top-down listening means using prior knowledge and experiences to understand (Brown, 2006).

Buck (2001) also stated that there are two types of information that help listeners determine what a word is. The first is parsing the acoustic signal in which clues such as pitch or stress help listeners to determine the phonemes and word boundaries. In second language listening, this type of information can present listeners with a source of difficulty to overcome because speech is a rapid, transitory medium that rarely relies on the isolation of individual words. Rather, individual words are grouped into idea units that take place over short stretches of time. Identifying these idea units is referred to as parsing. According to Buck, parsing idea units means determining the relationship between the parts of who does what, to whom, and with what, a process that is based on correctly interpreting semantic and syntactic clues. As Buck pointed out, in English some structures are more difficult to process than others. For example, negative statements take longer to process than affirmatives, passive statements take longer than active statements, or implausible events are harder to comprehend than plausible events. Idea units are hardest to process when both the semantic and syntactic cues conflict.

The second type of information that helps listeners determine what a word

is, is knowledge of the context, which concerns the listener's understanding of the situation in which the speaker is addressing. Listeners use phonological and suprasegmental knowledge as well as general background knowledge to segment the stream of sound into discrete words. Yi'an (1998) asserted that background knowledge played a role in how the participants answered the questions. For higher proficiency learners answering multiple-choice questions, background knowledge acted as a facilitator for the learner in that they could use the stem questions or distractors to make a decision. For lower proficiency learners, multiple-choice questions acted in a compensatory way to fill in any missing information.

Assessing Listening Comprehension

Rost (2011) pointed out that in addition to assessing listening ability, general language ability will also be assessed because listening ability is a subset of general language ability as seen in Figure 1. The figure also indicates how specific types of knowledge influences listening ability. General knowledge includes knowledge about the world, including the way people communicate. Pragmatic knowledge includes recognition of social dimensions in speech. Syntactic knowledge is based on the ability to parse speech at sentence and discourse levels. Lexical knowledge encompasses knowing the meaning of words and their relationship to other words and collocations. Phonological knowledge consists of knowledge of phonemes, allophonic variation, prosody, intonation, and stress. It also includes the application of this knowledge to recognize words in speech stream.

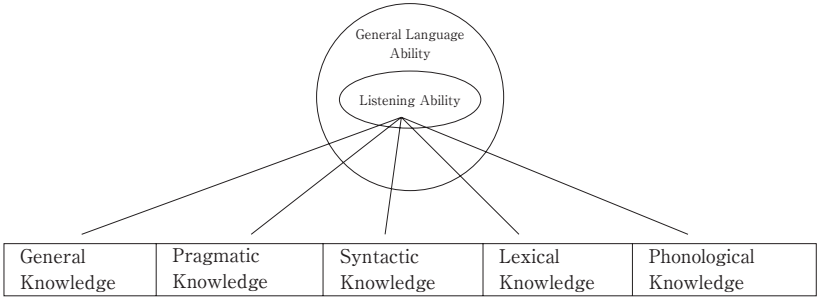

Figure 1. Model Integrating General language Ability and Listening Ability.

Test Factors

Numerous researchers have identified several factors influencing cognitive processing, which in turn influences test performance. Rubin (1994) and Rost (2002) identified and listed five such factors that should be considered in developing listening comprehension tests and assessing the level of cognitive processing involved. The first is text characteristics. The nature of the input stimuli influences listening comprehension in many ways, including speech rate, passage length, syntactic complexity, vocabulary, discourse structure, noise level, accent, register, propositional density, or amount of redundancy within the text. The second is text characteristics. The nature of the listening task can influence comprehension based on the visual context, the amount of context provided, clarity of instructions, response format, or the availability of question preview. The third is listener characteristics. Individual differences such as memory, interest, background knowledge, or motivation also influence listening comprehension. The fourth is interlocutor characteristics. For instance, the speaker's voice quality can influence comprehension, gender plays a role in influencing the stimuli pitch, and nationality affects pronunciation. The fifth is process characteristics. Listeners use bottom-up, top-down, and parallel processing to understand information. Although not entirely clear, there are a few conclusions drawn from the research: bottom-up processing taxes memory capacity more than top-down processing, and textual gist is identified prior to processing syntactic forms in top-down processing.

Bloomfield, Wayland, Rhoades, Blodgett, Linck and Ross (2011) examined factors affecting second language listening. They focused on the three characteristics: the listener, the listening passage, and the testing conditions. For the listener, usually those with greater working memory capacity can understand more of what they hear. Furthermore, those that use metacognitive strategies effectively demonstrate better listening comprehension. Additionally, as proficiency increases, the listener's ability to use bottom-up information correctly improves, whereas non-proficient users rely on background knowledge unsuccessfully to compensate for missing bottom-up information. For the listening passage, their conclusions are similar to Rubin (1994) and Rost (2002). They conclude that passage length is not as important as information density. Passage complexity is also a factor. Negative sentence structures, infrequent vocabulary usage, inferences, and pragmatic information such as idioms all decrease comprehension. The organization of the passage

affects comprehension. The greater the passage is organized with simplified syntax, clear discourse markers and redundant information, the more likely it makes listening comprehension easier. The testing condition is an area that can significantly affect comprehension even though the listener might be proficient. For example, by placing limits or controlling the time so that the listeners feel less in control can negatively affect scores. Generally, if given the chance, listeners will choose to replay passages. This may have a greater effect in reducing anxiety than in increasing comprehension.

The Spacing Effect

One of the most important aspects of a test is the retrieval of information from memory. Therefore the work of Ebbinghaus (1885/1964) is essential because of his tremendous influence in memory research. One of his most influential findings that has been confirmed repeatedly is the spacing effect. As he wrote, "with any considerable number of repetitions a suitable distribution of them over a space of time is decidedly more advantageous than a massing of them at a single time" (p. 89). Melton's (1967) research became the model of the spacing effect and lag effect. Participants saw a list of 48 words presented one at a time at a rate of one word every 1.5 seconds. Each word could appear once or twice. Words that were immediately repeated were called massed items. Repeated words that had intervening stimuli, i.e., other words, were spaced items. After the participants viewed the list, they were asked to recall the words. In a free-recall, words that had been repeated were recalled more often than words that had not been repeated. More importantly, spaced words were recalled more than massed words. An additional finding was that the greater the intervening stimuli length became between words the more the words were recalled, thus producing the lag effect. The spacing effect is the finding that memory performance is better when repetitions are separated by other items, i.e., spaced items, than when repetitions immediately follow one another, i.e., massed items. Massed items and spaced items receive the same total presentation time, but the spaced items are nonetheless recalled better than massed items.

Studies Addressing Repetition

Second language listening processes do not differ from first language listening processes in any physical aspect (Lynch, 1998) except that processing capacity is generally reduced (Call, 1985; Fishman, 1980; O'Malley, Chamot, &

Kupper, 1989; Voss, 1984). The difficulty for second language learners arises in comprehending specific elements of the language, and any necessary compensation, such as using background knowledge to modify the deficiencies, provides another opportunity for miscomprehension. Even two native speakers encounter misunderstandings and do not accurately comprehend everything they hear at all times. Several compensatory skills, such as using visual cues or common sense, can help compensate for incomplete listening comprehension.

Berne (1995) compared pre-listening tasks, repetition, and listening comprehension performance in which two of the three research questions were focused on repetition by examining the effect of multiple exposures to a listening passage on listening comprehension and the combined effect of pre-listening activities and multiple exposures to the listening passage. She tested 62 American university students studying Spanish in the third semester of a four-semester program. The participants were believed to be relatively experienced L2 learners of Spanish. Individual packets were randomly assigned to students in intact classes. The packets contained one of three types of pre-listening activities: a question preview activity, a vocabulary preview activity, and a filler activity for the control group. After completing one of the pre-listening tasks, the participants watched a video-taped lecture and answered 10 multiple-choice questions. Note-taking while watching the video was discouraged. The participants watched the same video again and took the same test again. The amount of time between watching the video was not stated, but at least several minutes transpired because the first test was collected and the new one was distributed.

A 2-way repeated-measures ANOVA was conducted. The dependent variable was the scores of two comprehension tests (test 1 and test 2), and the independent variables were pre-listening activities and exposure to the passage. The results were mixed for her first research hypothesis that pre-listening material would improve listening comprehension. The participants who previewed the test questions scored significantly higher than those who completed the filler activity, but those participants who previewed the test questions did not score significantly higher than those previewing vocabulary.

In regards to the second research hypothesis that repetition would increase comprehension, the results indicated that repeating the stimuli was beneficial to all participants. In her post hoc analysis, the contrasts revealed that the second comprehension scores (M = 4.36) (repeated listening passage) were significantly higher than the first comprehension scores (M = 3.33) for all pre-

listening activities (question preview activity, $p = .01$; vocabulary preview activity, $p < .01$; filler activity: $p = .01$). Berne argued, however, that a vocabulary preview as a pre-listening exercise might have distracted listeners' attention from the message content on the first listening presentation because of the small score differences between the two comprehension tests. This argument contradicted her results for the third research hypothesis.

In regards to the third research hypothesis, the lack of an interaction effect among the pre-listening activities, as noted above, indicated that repetition did not influence one pre-listening activity more than another.

Several limitations affected the results of this study. The test procedures were unusual if the participants had not been informed of the steps involved. The participants listened to the videotape and answered questions, the tests were collected and a new test was distributed, the participants listened again, and then took the same test. It was unclear if the participants knew the test procedures, particularly whether the listening input would be repeated and whether they knew that the same questions would be asked on the second test. In both cases, the interval between the tests might have caused enough disruption that their focus of attention could have inhibited recall. As Berne pointed out, there were several differences between the classroom material and test material. For example, the text used for the comprehension test was more difficult than the text used in class. Finally, none of the groups scored over five out of ten, which was less than 50% correct; this suggests that the listening passages or the tasks were too difficult for the participants.

Sherman (1997) examined the effect of question preview on listening comprehension tests with 78 undergraduate students majoring in social sciences at LUISS University in Rome, Italy. She used a counter-balanced Latin Square design to give four groups of students four different tests. All of the test formats included repetition. In the first format, students previewed ten short-answer questions, listened to the passage twice, and then answered the previewed questions. In the second format, students listened to the passage twice and then answered questions. In the third format, the students listened once to a text, read the questions, listened again, and then answered the questions. In the fourth format, students listened to the passage twice and then wrote down what they could remember of the story. In addition to taking the tests, the participants completed a questionnaire in which they expressed their opinions about each test format.

The results of an ANOVA, where the dependent variable was the test score

and the independent variable was the test version, indicated that the scores on the different tests did not differ significantly, but the third format had the highest mean score (50% compared to 42%, 42%, and 37%). Also, according to the questionnaire responses, 52 out of 135 participants favored the third format.

The results were a little surprising in that viewing the questions prior to the listening passage (format one) had the same results as viewing the questions after listening (format two). Sherman though speculated that the number of questions could have been too taxing and therefore negatively influenced recall and comprehension. The delayed repetition condition (format three) indicated positive results empirically and affectively.

Chang and Read (2006) investigated four listening support formats in which one of the conditions was repeated input. The first research question concerned whether different types of listening support would affect listening performance. The second research question asked whether the listening support types would affect higher or lower proficiency participants in the same manner. They examined the effects of the different formats on listening comprehension with 160 students from intact classes studying business at a college in Taipei, Taiwan. The participants, based on their class and class day, were given one test condition. Based on a TOEIC test, each group was further divided into low and high listening proficiency sub-groups. The participants in each condition completed two listening tests with 15 multiple-choice questions for each listening text. In the repeated input condition, the students were asked to listen to the text without any special preparation. Then they previewed questions before listening to the text twice, so they heard the text three times in all. Thereafter they answered 15 multiple-choice questions in three minutes. The steps were repeated for the second listening text.

Chang and Read conducted a 4 X 2 ANOVA. The dependent variable was the combined test score. The independent variables were four types of listening support (previewing questions, repeated input, topic preparation, and vocabulary instruction) and two listening proficiency levels (high and low). The results indicated that repeated input generated the second highest mean test scores.

For the first research question, significant main effects were found for listening support $F(3, 152) = 8.19$, $p < .05$ partial $\eta^2 = .13$, and listening proficiency $F(1, 152) = 8.31$, $p < .05$ partial $\eta^2 = .04$. There was also a statistically significant interaction between listening support and listening proficiency $F(3, 152) = 2.74$, $p < .05$ partial $\eta^2 = .04$. Two of the four types of

listening support, previewing questions (t = 3.53; p = .001) and repeated input (t = 2.84; p = .007), were statistically significant. These results indicated that the different types of listening support affected comprehension scores, but the effect sizes were small.

In the second research question, listening support activities affected the low and high proficiency levels differently. The main effects for listening support were statistically significant for both high proficiency students F(3, 152) = 6.20, p < .05, and low proficiency students F(3, 152) = 4.23, p < .05. The post hoc tests revealed that repeated input (t = 3.93; p = .000) and topic preparation (t = 3.17; p = .003) were significantly better than vocabulary input for high proficiency students. For low proficiency students, the results indicated that topic preparation (t = 3.17; p = .003) was significantly better than preview questions (t = 3.48; p = .001) and vocabulary input (t = 3.07; p = .004). These results indicated that high proficiency learners benefitted the most from repeated input, but the differences between two of the three listening support activities were not statistically significant. The low proficiency learners benefitted the most from topic preparation, but the difference of this activity from repeated input was not statistically significant whereas the other listening support activities were. Interestingly, although both levels of learners benefited from repeated input the most, the high proficiency learners were helped more than lower level learners.

The main drawback to this study was that the repeated input condition had both immediate and spaced repetition treatments. For example, the participants heard the passage twice (immediate treatment), were shown the questions, and heard the passage again (spaced treatment). The participants only viewed the questions between the second and third listening. It was not clear whether the participants were asked to actively answer the questions in their mind prior to the third listening. A second drawback was that another treatment condition could have been included to make comparisons between the repeated input condition and topic preparation condition. The repeated input condition did not have any topic preparation support, which influences schema building and thus listening comprehension. In this study, the first listening could be equated to a difficult form of topic preparation; and therefore, the results did not differ much from topic preparation.

Chang and Read (2007) examined listening support factors on listening comprehension with 140 students at a five-year post secondary educational program at a Taiwanese college with low levels of listening proficiency as

measured by the TOEIC listening section (a scaled score of 165 out of 465). Their primary research question that was investigated was what type of listening support, repetition, visual, or textual, would enhance comprehension for low-proficiency listeners.

A counter-balanced design was employed so that all participants took part in all three aspects of listening support. For the repetition support, students previewed the test items, listened once, and then responded to the test items. The students were then given a fresh test paper, they listened a second time, and responded to the test items again. There were three listening passages with 12 items per passage for a total of 36 multiple-choice and gap-filling items. The multiple-choice questions focused on general ideas whereas the gap-filling focused on specific information. The number of each question type was not stated.

To answer the first research question, a one-way ANOVA was conducted. The independent variables were visual support, textual support, and repeated support. The dependent variable was the listening comprehension score. The results indicated that all three types of listening support resulted in significantly higher scores than the control group $F(3, 556) = 20.16$, $p < .05$. Although no statistics were given, the authors conducted a post hoc test using Scheffé test that indicated that repeated input produced significantly higher scores than the other types of listening support.

There were several drawbacks in the study. The first was similar to the other studies in that repetition was only considered in one way, which in this case was spaced repetition. Additionally, the mean comprehension score in each group was about 50%, which suggests that the passages were either too difficult to comprehend and/or the task was too difficult.

Purpose of this Study

The primary purpose of this study is to examine the role of repetition using the spacing effect to compare two types of repeating stimuli input on a second language listening comprehension test. Thus, the independent variables are the control group, immediate repetition group, and spaced repetition group. In the immediate repetition condition, the input stimuli are repeated a second time prior to the participants answering listening comprehension questions. In this study, immediate repetition means the participants hear a passage twice before answering a set of five multiple-choice questions. Spaced repetition involves repeating the input stimuli after an intervening variable is given following the

first listening. Participants will listen to a stimulus, count from 1 to 5 out loud, hear the passage again, and then answer the five questions. The dependent variable is the test score from each condition.

A secondary purpose of this study was to determine if the different types of repetition affected question difficulty. For this study, question difficulty was considered in the following ways. First, Rost's (2011) model for integrating general language and listening ability was considered for developing certain types of questions. For example, phonological knowledge was used in making certain questions relating to Japanese difficulties distinguishing between /b/ and /v/ sounds. Second, Bloom's taxonomy (1956) was used as the basis as he identified six levels of cognitive difficulty. Brown (2001) interpreted Bloom's taxonomy for language purposes and outlined seven levels (p. 172). The first level and considered the easiest level was called knowledge questions. These types of questions ask for factual information, and test recall and recognition of information. The second level and considered more cognitively difficult is called comprehension questions. These types of questions ask for interpreting and inferring information. The fourth level and is also more difficult than the preceding levels is called inference questions. These questions include forming conclusions that are not directly stated. Third, Henning's (1991) definition was also adapted to make comparison more applicable. Therefore, lower-order cognitive difficulty was defined as questions requiring specific information stated in the passage within one sentence. Higher-order cognitive difficulty was defined as questions requiring information from two or more sentences or inferences.

Research Questions

1. Does the method of repetition affect listening comprehension scores for multiple-choice type questions?

2. Does the method of repetition affect question difficulty?

Hypothesis

1. Spaced repetition will increase comprehension scores more than immediate repetition.

2. There will be no significant difference between non-repetition and immediate repetition.

3. Spaced repetition will allow the participants to answer higher-order cognitive questions more accurately than immediate repetition.

METHOD

Participants

The students were enrolled at a coeducational national university in Japan. The school has a comprehensive TOEIC examination for all first year students. The participants were from seven intact classes of 240 students, majoring in Social Sciences. These students were placed in courses by their overall TOEIC placement score given at the beginning of the semester. Through self-admitted scores, most of the participants indicated that their TOEIC listening score was between 150-300 with their overall score between 300-500.

Materials and Design

The material used in this study consisted of 12 passages, six monologues and six dialogues. Each passage was followed by five multiple-choice questions. Each listening passage was approximately one minute in length based on the format of the listening passages used in the class. A Latin-square design with the passages and treatment conditions was arranged to limit any testing effects of procedures as seen in Table 1. The seventh class was not included in the design because it was used as a backup in case any assigned class had any problems. Since there were no problems, the seventh class followed the first class' design.

Table 1. Study Design

Class 1				
	Week one	Week two	Week three	Week four
Condition 1	Passage 1	Passage 4	Passage 7	Passage 10
Condition 2	Passage 2	Passage 5	Passage 8	Passage 11
Condition 3	Passage 3	Passage 6	Passage 9	Passage 12
Class 2				
	Week one	Week two	Week three	Week four
Condition 3	Passage 1	Passage 4	Passage 7	Passage 10
Condition 1	Passage 2	Passage 5	Passage 8	Passage 11
Condition 2	Passage 3	Passage 6	Passage 9	Passage 12

Class 3				
	Week one	Week two	Week three	Week four
Condition 2	Passage 1	Passage 4	Passage 7	Passage 10
Condition 3	Passage 2	Passage 5	Passage 8	Passage 11
Condition 1	Passage 3	Passage 6	Passage 9	Passage 12
Class 4				
	Week one	Week two	Week three	Week four
Condition 1	Passage 3	Passage 6	Passage 9	Passage 12
Condition 2	Passage 1	Passage 4	Passage 7	Passage 10
Condition 3	Passage 2	Passage 5	Passage 8	Passage 11
Class 5				
	Week one	Week two	Week three	Week four
Condition 3	Passage 3	Passage 6	Passage 9	Passage 12
Condition 1	Passage 1	Passage 4	Passage 7	Passage 10
Condition 2	Passage 2	Passage 5	Passage 8	Passage 11
Class 6				
	Week one	Week two	Week three	Week four
Condition 2	Passage 3	Passage 6	Passage 9	Passage 12
Condition 3	Passage 1	Passage 4	Passage 7	Passage 10
Condition 1	Passage 2	Passage 5	Passage 8	Passage 11

Note. Condition 1 = Control group; Condition 2 = Immediate repetition group; Condition 3 = Delayed repetition group.

Procedures

The experiment took place on the university campus over four class periods. The instructor used an overhead camera along with demonstration monitors to administer the test. The participants were told which procedure would be conducted prior to each listening passage. All the participants listened to the same listening passages. Prior to each listening section, the topic for each listening passage was shown on the monitor to help activate the test-takers' schema prior to listening. The students were able to take notes while listening.

Non-Repetition Test Procedures

In the non-repetition condition as shown in Table 2, the instructor gave verbal instructions while showing the English instructions, a sample question, and the listening passage topic on the overhead projector and computer demonstration monitors. Next, the listening passage was played via the classroom's audio system. The questions were not available for viewing while the listening passage was being played. After listening to the passage, the

participants were given one minute to answer five multiple-choice questions that were shown on the overhead projector and computer demonstration monitors. The instructor followed the same steps (instructions, playback, questions) for the remaining listening sections.

Immediate Repetition Test Procedures

During the immediate repetition condition as shown in Table 2, the procedures were as follows. First, the instructor gave verbal instructions in Japanese while showing the English instructions, a sample question, and the listening passage topic on the overhead projector and computer demonstration monitors. Next, the listening passage was played twice via the classroom's audio system. There was a short pause of five seconds between the first and second playback. The questions were not available for viewing while the listening passage was being played. After listening to the passage twice, the participants answered five multiple-choice questions. The questions were displayed via the overhead projector and computer demonstration monitors. The students were given one minute to answer five multiple-choice question items. The instructor followed the same steps (instructions, playback, questions) for the remaining listening sections.

Spaced Repetition Test Procedures

These were the procedures for the delayed repetition condition as shown in Table 2. First, the instructor gave the verbal instructions while showing the English instructions, a sample question, and the listening passage topic on the overhead projector and computer demonstration monitors. Next, the listening passage was played via the classroom's audio system. The questions were not available for viewing while the listening passage was being played. After listening to the passage once, the participants were asked to count, quietly but with vocal vibration, from 1 to 5 in order to disrupt working memory. The participants listened to the same passage a second time. Thereafter, the participants were given one minute to answer five multiple-choice questions that were shown on the overhead projector and computer demonstration monitors. The instructor followed the same steps (cover instructions, play the passage, participants count from 1 to 5, play the passage again, participants answer questions) for the remaining listening sections.

Table 2. Procedures for each condition.

Condition	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6
Control	Show topic	Play passage	Show questions	Answer questions		
Immediate repetition	Show topic	Play passage	Play passage again	Show questions	Answer questions	
Delayed repetition	Show topic	Play passage	Students count 5 to 1 (intervening variable)	Play passage again	Show questions	Answer questions

Scores

Individual scores were not used due to the design of the study. As the participants heard the passages under different conditions, a complete score for each participant in each condition was not possible. Rather, the scores were examined using the Rasch model. The item measure scores attained from Winsteps were used to conduct all analyses. As a reminder, the item measures in Rasch are used to determine difficulty so the higher the number, the more difficult the item is. It is hypothesized that spaced repetition will have higher scores than the control and immediate repetition conditions, so the item difficulty scores in the spaced condition should be lower than the scores in the control and immediate repetition conditions. Each condition had 12 passages with five multiple-choice questions, so there were 60 item difficulty measures for each condition.

RESULTS

Research Question 1: Does the type of repetition affect comprehension scores?

A one-way ANOVA was conducted to evaluate whether students scored higher on a listening comprehension test when the listening passages were repeated under different conditions, i.e., control, immediate or spaced. The assumptions were checked and met for conducting the ANOVA. The descriptive statistics are shown in Table 3. Three items were deleted because of z-score values greater than ± 3.29. An additional three items were deleted because boxplots indicated they were outliers for the treatment condition or question difficulty. The results indicated that the mean score for the spaced repetition condition (M = 49.90 SD = 2.80) was not significantly higher than the

mean score for the immediate repetition condition (M = 49.85 SD = 2.96), or control condition (M = 50.61 SD = 2.65), $F(2, 168) = 1.37$, $p > .05$.

Table 3. Descriptive Statistics for each Condition on the Listening Test.

	Control	Immediate Repetition	Spaced Repetition
N	58	59	57
M	50.61	49.85	49.90
SE	0.34	0.38	0.36
SD	2.65	2.96	2.80
Skewness	0.27	-0.24	0.22
SES	0.31	0.31	0.32
Kurtosis	-0.45	-0.44	-0.11
SEK	0.62	0.61	0.62

Figure 2 shows the boxplots for the three treatment conditions. They indicate that the median for each condition is almost equal with one another. Additionally, the range for each condition is dispersed fairly equally with the immediate repetition condition having the greatest range.

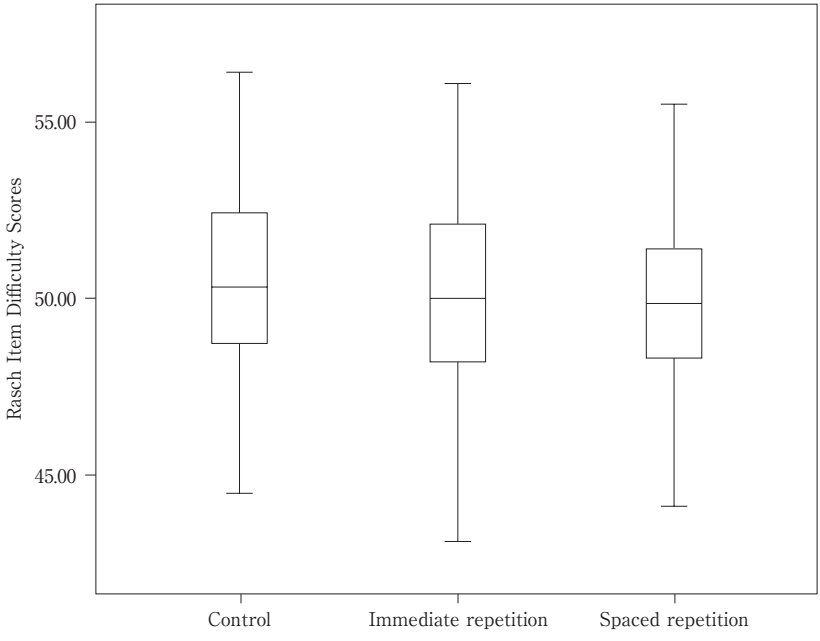

Figure 2. Boxplots of the Rasch item difficulty scores for each condition.

Research Question 2: Does the method of repetition affect different types of question difficulty?

A 3 x 2 ANOVA was conducted to evaluate the effects of three conditions of repetition (control, immediate, and spaced) and question difficulty (low and high) on listening comprehension scores. The assumptions were checked and met for conducting the ANOVA. The descriptive statistics are presented in Table 4. The ANOVA indicated no significant interaction between repetition and question difficulty, $F(2, 168) = 1.39$, $p > .05$. No further analysis was conducted.

Table 4. Descriptive Statistics for Question Difficulty in each Condition.

Condition	Control		Immediate Repetition		Spaced Repetition	
Question difficulty	Low	High	Low	High	Low	High
n	33	25	34	25	33	24
M	50.49	50.76	49.87	49.8	50.78	49.72
SE	0.48	0.55	0.47	0.55	0.48	0.55
SD	2.71	2.51	2.87	3.14	2.44	2.61

DISCUSSION

Research Question 1 asked whether the method of repetition affected the listening comprehension scores. The first hypothesis stated that spaced repetition would aid listening comprehension more than immediate repetition. As the results indicated no significant effect between the conditions, the hypothesis was not met. The second hypothesis stated there would not be a difference between the control condition and immediate repetition. There was no statistical distinction between the three conditions; neither type of repetition had any statistical benefit. In this case the hypothesis was met. By examining Figure 3, however, it is clear that both types of repetition were better than control condition of listening once. Although the listening comprehension scores were slightly greater in the repetition conditions, overall, the method of repetition had little effect on the comprehension scores. This finding was similar to Sherman (1997) in that the scores were higher for repetition, but not significantly higher statistically.

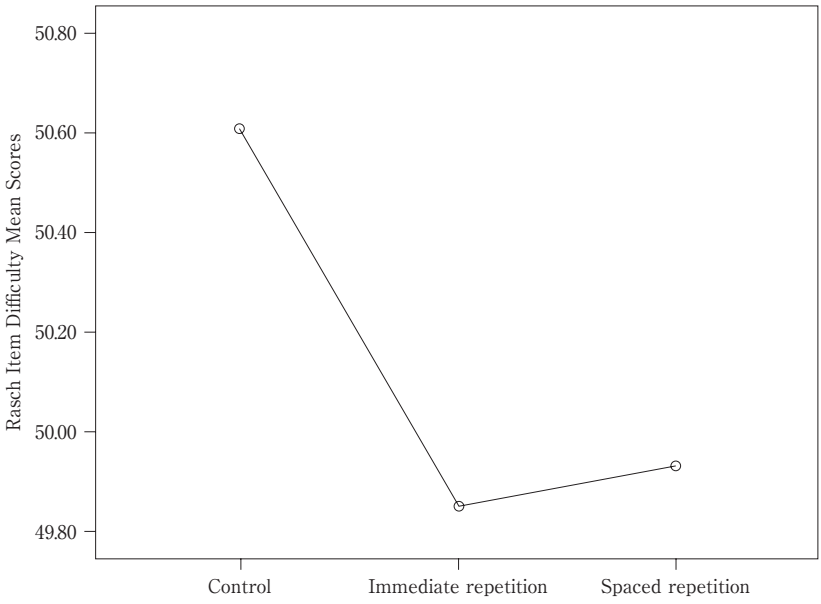

Figure 3. Mean scores for each condition based on the Rasch item difficulty measures.

Research question 2 asked whether the method of repetition affected question difficulty. The third hypothesis was that spaced repetition would allow the participants to answer higher-order cognitive questions more accurately than immediate repetition. There were no statistical differences between the two types of question difficulty. The results indicate that repetition does not affect question difficulty. The mean score for each condition and question type was nearly equal. As the hypothesis was not met, it seems spaced repetition neither useful or useless based on question difficulty.

Although the results were not supportive of the hypothesis, there are factors to be considered. First, the listening test questions were slightly above the participants' listening ability as indicated by the Rasch scores (see Appendix A, figure A1, person-item map). Second, the participants' listening proficiency range was narrow so many of the questions were redundant thus limiting any effect. Finally, due to limitations of the scan mark-sheet and participants' notes not collected, it is unclear whether the participants took notes effectively. Some mark-sheets had notes on them while others did not. These factors could have made it so that the participants were unable to use background knowledge to compensate for missing information, thus the lack of distinction between lower- and higher-order question difficulty. In addition, as Bloomfield et

al. (2011) suggested that using meta-cognitive increases comprehension, perhaps issuing notepaper and requiring notes would have encouraged better note-taking, and thus increase the likelihood of better comprehension scores.

The results in this study indicate that repetition did not effectively increase comprehension scores using multiple-choice items and varying question difficulty. Although the Rasch item measures indicate that repetition makes items easier, the difference is not significant.

REFERENCES

Berne, J. E. (1995). How does varying pre-listening activities affect second language listening comprehension? *Hispania, 78*, 316-319.

Bloom, B. (1956). *Taxonomy of educational objectives, handbook I: The cognitive domain*. New York: David McKay Co. Inc.

Bloomfield, A., Wayland, S. C., Rhoades, E., Blodgett, A., Linck, J., Ross, S. (2011). What makes listening difficult? Factors affecting second language listening comprehension. Technical Report TTO 81434 E.3.1, University of Maryland Center for Advanced Study of Language, College Park, MD.

Brown, H. D. (2001). *Teaching by principles: An interactive approach to language pedagogy*. White Plains, NY: Longman.

Brown, S. (2006). *Teaching listening*. Cambridge: Cambridge University Press.

Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.

Call, M. E. (1985). Auditory short-term memory, listening comprehension, and the

Chang, A., & Read, J. (2006). The effects of listening support on the listening performance of EFL learners. *TESOL Quarterly, 40*, 375-397.

Chang, A., & Read, J. (2007). Support for foreign language listeners: its effectiveness and limitations. *RELC Journal 38*, 375-395.

Ebbinghaus, H. (1964). Memory: A contribution to experimental psychology, (H.A. Ruger and C.E. Bussenius, Trans.). Dover, New York. (Original work published 1885).

Fishman, M. (1980). We all make the same mistakes: a comparative study of native and non-native errors in taking dictation. In J. Oller & K. Perkins (Eds.) *Research in Language Testing* (pp.184-194). Rowley, MA: Newbury House.

Henning, G. (1991). A study of the effects of variation of short-term memory load, reading response length, and processing hierarchy on TOEFL listening comprehension item performance. Retrieved from NJ: Educational Testing Service. Website: http://www.ets.org/toefl/research/archives/research_report.

Lynch, T. (1998) Theoretical perspectives on listening. *Annual Review of Applied Linguistics, 18*, 4-19.

Melton, A. W. (1967). Repetition and retrieval from memory. *Science, 158*, 532.

O'Malley, J.M., Chamot, A.U., & Kupper, L. (1989). Listening comprehension strategies in second language acquisition. *Applied Linguistics 10*, 418-437.

Rost, M. (2002). Teaching and researching listening. London: Longman.

Rost, M. (2011). *Teaching and Researching Listening* (2nd Ed.). London: Pearson Education

limited.

Rubin, J. (1994). A review of second language listening comprehension research. *Modern Language Journal, 78*, 199-221.

Sherman, J. (1997). The effect of question preview in listening comprehension tests. *Language Testing, 14*, 185-213.

Voss, B. (1984). *Slips of the Ear*. Tubingen: Gunter Narr Verlag.

Yi'an, W. (1998). What do tests of listening comprehension test?—A retrospection study of EFL test-takers performing a multiple-choice task. *Language Testing, 15*, 21-44.

Appendix A

Figure A-1 shows the Wright map, which is the person-item relationship in a pictorial representation (Bond & Fox, 2007). The CHIPS scale is shown on the far left side of the figure. According to Linacre (2008) CHIPS are a useful transformation, in which 1 logit = 4.55 CHIPS. In this user-scaling system, standard errors tend to be about 1 CHIP in size. A comparison of the locations of the person measures (left side) and item measures (right side) shows that the mean CHIPS for the person measures for the participants (M = 47.2; SD = 2.2) is equal to the mean of the item measures (M = 50.00; SD = 3.2). In addition, Figure 1 shows many of the item measures were redundant in that they shared the same location on the logit scale with at least one other item. Nonetheless, the item measures were spread out sufficiently in that the items range were beyond the participants' listening comprehension ability on this test. Along the left side of the map, the participants are spread out over 15 CHIPS, minimum = 41.4 CHIPS, maximum = 56.9 CHIPS, with higher proficiency students toward the top of the map and lower proficiency students toward the bottom. Along the right side of the map, the items spread out over 23 CHIPS, minimum = 37.7 CHIPS, maximum = 62.8 CHIPS, with the easier items toward the bottom of the map and the more difficult items toward the top. In this case, the listening comprehension test covers the abilities of the lowest to the highest students so there are no ceiling or floor effects. The common linear interval data for persons and items gives a clear demonstration of whether the items matched the persons' abilities for the construct measured. The items are above the student's ability on average as the means in the middle of the map indicate the items were above the students' ability.

More able persons		|	More difficult items		
		|	Item 16-L-C		
60		+			
	.	|			
	.	|T	Item 36-H-C	Item 36-H-IR	Item 38-L-C
			Item 43-H-SR	Item 50-H-IR	
		|	Item 10-H-C	Item 16-L-SR	Item 2-L-SR
			Item 3-H-SR	Item 36-H-SR	Item 43-H-C
			Item 49-L-C	Item 52-L-C	Item 58-L-C
			Item 58-L-SR	Item 6-L-C	Item 6-L-SR
			Item 8-H-C		
	.	|S	Item 16-L-IR	Item 2-L-IR	Item 20-H-C
			Item 20-H-SR	Item 24-H-C	Item 25-L-C
			Item 3-H-C	Item 3-H-IR	Item 30-H-C
			Item 33-L-C	Item 38-L-SR	Item 38-L-IR
			Item 41-L-SR	Item 43-H-IR	Item 54-L-IR
			Item 58-L-IR	Item 6-L-IR	
	.# T	|	Item 10-H-SR	Item 11-L-IR	Item 12-H-IR
			Item 13-H-SR	Item 13-H-IR	Item 15-L-IR
			Item 19-H-SR	Item 19-H-IR	Item 20-H-IR
			Item 23-L-C	Item 28-L-SR	Item 28-L-IR
			Item 30-H-IR	Item 33-L-SR	Item 39-L-C
			Item 39-L-SR	Item 39-L-IR	Item 41-L-C
			Item 41-L-IR	Item 42-L-SR	Item 44-L-C
			Item 44-L-SR	Item 44-L-IR	Item 50-H-C
			Item 50-H-SR	Item 52-L-SR	Item 52-L-IR
			Item 55-L-C	Item 57-L-IR	Item 59-L-C
			Item 59-L-IR	Item 8-H-SR	Item 8-H-IR
50	#####	S+M	Item 10-H-IR	Item 11-L-SR	Item 12-H-SR
			Item 14-L-IR	Item 15-L-C	Item 17-H-C
			Item 17-H-SR	Item 19-H-C	Item 2-L-C
			Item 22-H-C	Item 22-H-SR	Item 23-L-IR
			Item 25-L-SR	Item 25-L-IR	Item 26-H-IR
			Item 27-L-C	Item 27-L-IR	Item 28-L-C
			Item 30-H-SR	Item 32-H-C	Item 33-L-IR
			Item 34-L-SR	Item 42-L-C	Item 47-L-C
			Item 47-L-SR	Item 49-L-SR	Item 49-L-IR
			Item 5-L-SR	Item 54-L-C	Item 54-L-SR
			Item 55-L-IR	Item 56-H-C	Item 57-L-C
			Item 59-L-SR	Item 60-H-C	Item 60-H-SR
			Item 60-H-IR	Item 7-L-C	Item 7-L-SR
			Item 7-L-IR		
	.########	|	Item 11-L-C	Item 12-H-C	Item 13-H-C
			Item 14-L-SR	Item 15-L-SR	Item 17-H-IR
			Item 18-L-C	Item 22-H-IR	Item 23-L-SR
			Item 24-H-SR	Item 24-H-IR	Item 26-H-C
			Item 27-L-SR	Item 29-H-C	Item 29-H-SR
			Item 31-H-C	Item 31-H-SR	Item 32-H-SR
			Item 34-L-C	Item 34-L-IR	Item 35-L-C
			Item 37-H-C	Item 40-L-C	Item 40-L-SR
			Item 40-L-IR	Item 42-L-IR	Item 45-L-C
			Item 46-L-C	Item 46-L-SR	Item 47-L-IR
			Item 48-H-C	Item 48-H-SR	Item 5-L-C
			Item 5-L-IR	Item 53-H-C	Item 53-H-SR
			Item 53-H-IR	Item 55-L-SR	Item 56-H-SR
			Item 56-H-IR	Item 57-L-SR	Item 9-H-SR
	.###########	M|S	Item 14-L-C	Item 18-L-SR	Item 21-L-C
			Item 21-L-SR	Item 29-H-IR	Item 31-H-IR
			Item 32-H-IR	Item 35-L-SR	Item 37-H-SR
			Item 46-L-IR	Item 48-H-IR	Item 51-H-C
			Item 51-H-SR	Item 9-H-C	
	#########	S|	Item 1-L-C	Item 18-L-IR	Item 35-L-IR
			Item 37-H-IR	Item 45-L-SR	Item 45-L-IR
			Item 51-H-IR	Item 9-H-IR	
	##	T|T	Item 1-L-IR	Item 21-L-IR	Item 26-H-SR
	.	|	Item 1-L-SR		
40		+	Item 4-H-SR		
		|	Item 4-H-C		
		|	Item 4-H-IR		
30		+			
Less able persons		|	Less difficult items		

Note. Each #' = 7 people; each . = 1 - 6 person; L = Specific detail question type; H = inference question type; C = Control condition; IR = Immediate repetition condition; SR = Spaced repetition condition.

Figure A-1. Person-item map for the pilot listening comprehension test.