International Journal of Research Studies in Language Learning
2017 January, Volume 6 Number 1, 95-116

Raising awareness of unstressed words for dictation accuracy: A study on the effects of task and Japanese EFL learners' proficiency and attitude

Ofuka, Etsuko ✉
Kobe Pharmaceutical University, Kobe, Hyogo, Japan (ofuka@st.kobepharma-u.ac.jp)
Gilbert, Joan E.
Kwansei Gakuin University, Nishinomiya, Hyogo, Japan (z98325@kwansei.ac.jp)

ISSN: 2243-7754
Online ISSN: 2243-7762

OPEN ACCESS

Received: 24 October 2015 **Revised**: 22 November 2015 **Accepted**: 2 February 2016
Available Online: 16 February 2016 **DOI**: 10.5861/ijrsll.2016.1352

Abstract

This study investigates whether or not an intervention task improved the dictation accuracy of 578 Japanese college students with different majors and English proficiency levels. The task consisted of (a) cloze dictation, focusing on unstressed functional elements; (b) collaborative group completion of cloze dictation activities; and (c) output, which was shadowing of the dictation sentences. The study used one control group and two experimental groups. One experimental group did the activities of (a) and (b), while the other experimental group did all three activities. The intervention effect was assessed by recognition of unstressed functional elements (UF) through pre-post dictation tests. Statistically significant results were noted in the dictation accuracy progress for the experimental groups, but the effectiveness of the output activity was inconclusive. The participants' fields of study and proficiency levels, which were determined by their pretest scores, were also found to influence their post-dictation performance. The lowest English-level participants made the most progress but were unable to make the same progress on UF items that upper-level participants did. The effect of the intervention was also assessed by participants' self-reported strategic use of paying attention to UF (AUF) through pre-post questionnaires, but the intervention was not significantly related to changes in AUF scores. It was found that pre-AUF scores were related to pre-dictation performance, and participants with higher pre-AUF scores demonstrated higher task performance. Educators could use this type of intervention, which could be implemented easily in any type of classroom, to improve EFL learners' dictation accuracy.

Keywords: cloze dictation; collaborative group discussions; learners' readiness; field of study; learning strategy

Raising awareness of unstressed words for dictation accuracy: A study on the effects of task and Japanese EFL learners' proficiency and attitude

1. Introduction

1.1 Background

Global competition in business, science, technology, and education sectors has created a need for non-native speakers of English to acquire communicative competence. Without accurate language production, miscommunication or communication breakdowns during speech acts involving non-native speakers of English or non-native and native speakers of English could occur. According to Field (2003), even simple listening recognition errors (e.g., hearing *I won't go to London* as *I want to go to London*) could cause serious communication breakdowns (p. 325).

The importance of listening is vital for any learner of a language. However, simple exposure to a target language does not seem to be enough for adult learners to achieve near-native accuracy (e.g., Doughty & Williams, 1998b). Even for young children, studies have shown that long exposure or input flooding of a target language is insufficient for accurate language production. In a review of six studies done on the immersion approach, Hammerly (1987) claimed that the results of the studies demonstrated that even after several thousand hours of input flooding of the target language, the young participants did not improve beyond "terminal classroom pidgin" ability (p. 397).

The problem with input flooding is that learners do not tend to pay attention to specific features of the target language, which is needed for accurate language production. Attention is considered necessary for any learning process for adults (Baars & Gage, 2012). Schmidt (1990) emphasized the importance of learners' noticing and awareness in adult language learning and stated that "noticing is the necessary and sufficient condition for converting input to intake" (p. 129). This could imply that the more attention adult learners pay to the target language features the more learning that occurs, and if noticing is a prerequisite for adult learning as Schmidt claimed, then corrective feedback would be necessary. However, research has shown that corrective feedback is mostly factual or discourse-related and does not focus on grammatical errors.

In a study done by Chun, Day, Chenoweth, and Luppescu (1982), less than 10% of the grammatical errors produced by non-native speakers were corrected by native speakers in social contexts. Even in classroom situations, language teachers' corrections of grammatical errors went largely unnoticed by language learners (e.g., Lasagabaster & Sierra, 2005; Roberts, 1995; Slimani, 1992). This could imply that intentional awareness-raising of grammatical features may be necessary for adult language learners. In a recent article, DiCerbo, Anstrom, Baker, and Rivera (2014), stated that renewed attention on instruction and feedback of grammatical features for language learners is being supported, and instead of explicit instruction, focusing on grammatical forms along with classroom discourse, including student interactions in groups, is needed. Doughty and Williams (1998a) claimed that learners' awareness of target features could be increased through an intervention approach termed *focus on form* in which learners focus on linguistic features while participating in a communicative activity. Doughty and Williams (1998b) also stated that collaborative group discussions helped students focus on linguistic features through the process of reflecting and discussing the target linguistic forms, and the effectiveness of collaborative learning has been documented (e.g., Crouch & Mazur, 2001; Jones, Antonenko, & Greenwood, 2012; Abadikhah & Shahriyarpour, 2012).

Which linguistic features need to be focused on varies according to the language learners. For language learners of English, unstressed functional elements are difficult because they are nonsalient informationally and

perceptually (e.g., Field, 2003). In the case of Japanese English-language learners, they find noticing unstressed functional elements, for example, unstressed word-endings, articles, and function words, extremely difficult. In a pilot study done by Ofuka (2015), over 50% of the students (n = 82), who were at the high beginning to low intermediate level, were unable to notice their unstressed function word errors and failed to make the corrections for a dictation task (16 sentences with an average of 7 words each) even with the script in hand. The difficulty of noticing unstressed function-word errors might stem from the differences in Japanese and English. Unlike English, Japanese has no weak forms. Japanese also has mostly non-closed syllables, so Japanese speakers do not naturally pay attention to the final consonant when listening to English, which could cause them to miss vital information concerning tense or comparatives. Based on prior studies showing that language learners find it difficult to notice unstressed function elements and error correction was found to be largely ineffective, the present study focused on tasks that could effectively draw learners' attention to unstressed functional elements.

In addition to task effects, prior research has shown that there are various factors that could influence learners' task performance (Dörnyei, 2006; Zafar & Meenakshi, 2012), for example, learners' proficiency levels and attitudes toward tasks. Williams and Evans (1998) found in their study that learners' ability to focus on a certain linguistic feature during an intervention task was related to their eventual mastery of that feature. For the attitudinal factor, in a study involving an oral task, in which participants were assessed by the number of words spoken and turns taken, the participants that had positive attitudes toward the English course and intervention task were able to do significantly better (Dörnyei & Kormos, 2000), and in another study, the participants were able to produce more accurate outputs in English if they were motivated and had positive attitudes toward the English course (Kormos & Dörnyei, 2004). Based on these findings, the authors hypothesized that the participants' attitudes toward English would also have an impact on their performance in this study. We assumed that the participants' attitudes toward English might be related to their field of study and their perceived need for English for their future, so we used their field of study as the motivational/attitudinal factor in our analyses. In this study, we refer to *field of study* as science and technology (S&T) majors or humanities and business (H&B) majors. For S&T majors, their primary focus is on passing national licensing examinations, which do not include any English; whereas, H&B majors might need English for the national teaching license or business encounters in globalizing societies. In addition, the authors assumed that the difference in the instructors' nationality (Japanese for S&T and American for H&B) might affect the participants' motivation to use English.

1.2 Objectives and research questions

This study employed an awareness-raising task designed for Japanese EFL university learners to help them pay more attention to the unstressed part of spoken input. It is a partial replication of Morris and Tremblay's (2002) study, which investigated the effect of an intervention on French-speaking ESL junior college students' acquisition of unstressed grammatical words. The intervention included the following: (a) a cloze dictation activity designed to focus participants' attention on unstressed grammatical words, (b) a group work activity to discover errors or missing words in their answers, and (c) a second cloze activity, in which students had to fill-in every seventh word. For this study, the first two activities of Morris and Tremblay's intervention were used, and shadowing was added to the cloze dictation activity for one of the experimental groups for the purpose of examining whether an added productive intervention would have an influence or not.

This study is unique in three main ways. First, in Morris and Tremblay's (2002) study, the outcome was assessed by free story telling output; whereas, in this study, the effects were assessed by participants' ability to recognize unstressed grammatical elements through pre-post dictation tests and their self-reported awareness levels of unstressed elements. To examine the awareness-level change, we asked the participants to indicate to what extent they paid attention to unstressed elements while listening to English in both the pre- and post-questionnaires. These data were used to examine whether or not participants' participation in the intervention activities would influence their strategic use of paying attention to unstressed elements. Language learning requires long-term effort, for example remembering sound changes due to coarticulation in connected speech and appropriate use of prepositions. Therefore, we assumed language learners need to be able to employ

awareness strategies whenever listening to English in order to facilitate long-term language learning. Second, in Morris and Tremblay's study, the time spent was about one hour for the activities per week compared to only 15-20 minutes per week in this study because we wanted to encourage active participation and facilitate transferability of this intervention. Finally, Morris and Tremblay's study was done in Canada with participants who were French-speaking ESL learners and at similar proficiency levels, and the control groups and experimental groups were taught by different instructors; however, in this study, which was carried out in Japan, the sample was large with 578 Japanese EFL learners, who were at mixed proficiency levels and had different fields of study, and the instructors taught both control groups and experimental groups to control instructor influence (Guilloteaux & Dörnyei, 2008).

Our research questions are summarized as follows:

- To what extent did the awareness-building and collaborative intervention influence participants' dictation performance of unstressed elements?
- Did the intervention tasks have an influence on participants' self-reported strategic use of paying attention to unstressed elements?
- Did any of these two factors (proficiency levels, and learners' attitudes toward English and task) influence participants' (a) post-performance of unstressed elements and (b) self-reported strategic use of paying attention to unstressed elements?

2. Materials and Methods

2.1 Pre-Post Assessments

This is a quantitative study that employed a pre-posttest design with assessments consisting of questionnaires and dictation assessment tests.

Questionnaires - The pre-questionnaire was administered before the start of the study as a data-eliciting tool to obtain information on the participants' English language background and attitudes toward English learning (see Appendix A). The questions included: (a) at what age and how long they had lived in another country, (b) what languages were used in their daily life, (c) attitudes toward English learning, and (d) to what extent they pay attention to intonation and unstressed words while listening to English.

The post-questionnaire, containing scale items, was administered at the end of the study to collect information on the extent to which participants paid attention to intonation and unstressed words when listening to English. The participants were asked to respond to each scale item by marking any place on the scale, a sample of which is shown in Figure 1. For the experimental groups, the questionnaire included extra scale items with which the participants were asked to provide their reflections of the intervention tasks, which are shown in Appendix B. In order to ensure the participants understood the questions and instructions, they were written in Japanese.

1. 英語を聞くとき、英語の弱く話される箇所に注意を向けますか。(以下に ｜ を入れてください)
(When you listen to English, do you pay attention to unstressed words?) (Mark on the scale.)

(not at all) (somewhat) (very much)
考えたこともない 少しは注意する とても注意して聞く
|----------|----------|

Figure 1. A sample of the rating scales in the pre-post questionnaires

Dictation assessment test - The same dictation test was administered in the first week (pretest) and the second to last week (posttest) of a 15-week semester. The listening material consisted of four TOEIC® Part-II-type exercises. Before the initial dictation test, the instructors explained the pronunciation and meaning of eight words that were considered unfamiliar to some lower-level students. Then, the students listened to an audio-recording of the material once for meaning. Next, they heard each of the four sections three times. The instructors stopped the audio-recording after each sentence and gave the participants time to write the sentence down as accurately as they could on the designated dictation paper. Immediately after the test was completed, instructors collected the dictation papers. The entire process took about 15 minutes.

2.2 Session procedures

There were one control group and two experimental groups (*Dg* and *DgS*). All participants in experimental groups had seven intervention sessions. All seven intervention sessions were conducted at the beginning of each class by the class instructor.

Before the first task, the participants in the experimental groups were divided into small groups of 3-5 members either by lottery or student ID numbers. The procedures are in Table 1. The task consisted of three activities: (a) cloze dictation, (b) group discussion (Step 5), and (c) output exercise, which was shadowing the sentences used in the dictation (Steps 9-11). The Dg group was assigned (a) and (b), while the DgS group did (a), (b), and (c). For the DgS group, the instructor explained the method and objective of the shadowing exercises in the second week. Each session took 15-20 minutes. The control group did not receive any intervention task.

Table 1

Session Procedures

Step	Procedures
1	Play the material once for meaning.
2	Distribute the answer sheet.
3	Play each sentence once. Participants write down an appropriate word in parentheses.
4	Play each sentence for the second time.
5	Distribute the answer sheet for groups. Participants discuss the answers with group members and fill in the group answer sheet.
6	Play each sentence for the third time. The participants decide on the final version of the group answer sheet.
7	Collect the group answer sheet.
8	Write down the correct answers on the board. The participants check the answers.
The following additional steps are for the DgS group.	
9	Play each sentence once. Participants listen carefully and silently repeat with the script.
10	Play each sentence for the second time. Participants do the same as Step 9, being encouraged not to look at the script.
11	Play each sentence for the third time. Participants do the same as Step 9, using a whisper-like voice (shadowing) without the script.

The seven cloze dictation intervention tasks had a combined 101 spaces indicated by parentheses that the participants were to fill with pronouns (28.7%), prepositions (20.8%), articles (18.8%), word-final inflections (16.7%), auxiliary verbs (7.9%), and contracted forms (6.9%). A sample of the tasks is in Appendix C.

2.3 Participants

This study included a convenience sample of 578 Japanese EFL students in their first or second year at four universities in Japan, and they belonged to two types of classes (S&T and H&B) (see Table 2 for description). All participants ranged from 18 to 20 in age and had studied English as a compulsory subject for six years before

entering university. A total of 153 participants' data (20.9%) were excluded from all the analyses for the following reasons: (a) participants who lived abroad for over 3 months, (b) those who used a foreign language in daily life, (c) those who did not participate in either the pretest or posttest, and (d) those who did the same pretest in the previous year (21 students). The male ratios in the sample population were similar: 38.6% for S&T classes and 37.5% for H&B classes. Data were collected during six semesters between April 2012 and December 2014. The summary of groups of participants and assigned tasks are in Table 3.

Table 2

Description of Class Types

S&T classes	H&B classes
science and technology majors	literature, education & business majors
9 classes at 2 colleges (n = 266)	19 classes at 2 colleges (n = 465)
(average 30.7 students/class)	(average 23.7 students/class)
a female instructor	a female instructor
who was a native speaker of Japanese	who was a native speaker of American English
Objective was listening & speaking.	Objective was listening & speaking.
(Supplementary dictation exercises	(No supplementary dictation exercises
with key word focus done in class)	done in class)

Table 3

Summary of Groups of Participants and Assigned Tasks

Assigned task	Control group		Experimental group			
			Dg[a]	Dg	DgS[b]	DgS
Class type	S&T	H&B	S&T	H&B	S&T	H&B
# of classes x university (year)	3 x U1(2^{nd})[d] 2 x U2(2^{nd})	4 x U3(1^{st})[d] 1 x U4(1^{st}) 1 x U3(2^{nd}) 1 x U4(2^{nd})	2 x U1(2^{nd})	3 x U3(1^{st}) 1 x U4(1^{st}) 2 x U3(2^{nd})	2 x U1(2^{nd})	3 x U3(1^{st}) 1 x U4(1^{st}) 2 x U3(2^{nd})
enrolled[c]	91 + (21)[e]	164	77	138	77	163
sample size[e]	80	138	56	112	61	131

Note. [a]Dg is a group of participants who did dictation and group work;
[b]DgS is a group of participants who did dictation, group work, and shadowing;
[c]enrolled is the number of enrolled students; [d]sample size is the number of students whose data were used for analyses;
[d](1^{st}) is 1^{st}-year students, and (2^{nd}) is 2^{nd}-year students; and
[e](21) is the number of students who did the same pretest in the previous year.

3. Results

3.1 Dictation performance of unstressed functional elements and effects of related factors

We used data from the participants' performance on 33 unstressed functional elements (UF) items, which consisted of unstressed functional words, contracted elements, and morphological elements (see Table 4), on the pretest and posttest. Participants' performance of UF (UF dictation accuracy rate) was calculated by the number of UF items dictated correctly divided by the total number of UF items (= 33). UF dictation progress was measured by subtracting the UF accuracy rate on the pretest from that on the posttest. As was stated in the third research question, participants' proficiency levels and class types in addition to the task factor were included in the analyses.

Table 4

UF Items Included in Pretest and Posttest

UF item	UF item in context with the feature in ( )	Grammatical feature
UF01:	(Do) you think …	auxiliary verb
UF02:	... stole (my) wallet.	possessive
UF03:	… should hide (it) now.	personal pronoun
UF04:	No, (my) bag …	possessive
UF05:	... my bag (was) not stolen.	auxiliary verb
UF06:	I believe he'(s) the last man …	contraction
UF07:	… he's (the) last man …	article
UF08:	… the last man (to) do …	preposition
UF09:	… such (a) thing.	article
UF10:	Why (were) you late …	be-verb
UF11:	… late (for) the meeting.	preposition
UF12:	… late for (the) meeting.	article
UF13:	I guess I (was) 10 minutes late.	be-verb
UF14:	… 10 minute(s) late.	plural
UF15:	I (was) caught …	auxiliary verb
UF16:	… caught (in) a heavy traffic jam.	preposition
UF17:	… caught in (a) heavy traffic jam.	article
UF18:	Who (is) the new class monitor?	be-verb
UF19:	Who is (the) new class monitor?	article
UF20:	Yes, (the) new class monitor …	article
UF21:	… monitor (will) be elected …	auxiliary verb
UF22:	… monitor will (be) elected…	be-verb
UF23:	… will be elect(ed) soon.	past participle
UF24:	I think John (is) the one elected.	be-verb
UF25:	… John is (the) one elected.	article
UF26:	… the (one) elected.	indefinite personal pronoun
UF27:	… is the one elect(ed).	past participle
UF28:	… what (do) you want …	auxiliary verb
UF29:	… want (to) do tonight?	preposition
UF30:	… rent (a) DVD?	article
UF31:	I'(m) not going …	contraction
UF32:	He want(s) me to do …	third person singular
UF33:	… wants me (to) do that soon.	preposition

Note. The labels of the grammatical features were based on Swan (2005).

Task factor - The task factor consisted of *control* (CTL) and *experimental* (EXP). The EXP participants were divided into two groups and experienced slightly different interventions as follows: Group *Dg* did dictation tasks followed by group work; and Group *DgS* did dictation tasks, group work, and shadowing (see 2.2 Session procedures for more details).

Factor related to proficiency levels - Participants' proficiency levels were determined based on their overall dictation score, which was calculated as the number of correctly dictated items (words and contracted/morphological elements) divided by the total number; the contracted and morphological elements (i.e., UF06, UF14, UF23, UF27, and UF32 in Table 4) were defined as separate items. Each word on the pretest and posttest that was misspelled (e.g., *minits* for *minutes*) was counted as 0.5 points instead of 0 points. The unfamiliar words introduced before the pretest, and the words (e.g., *Yes*, *No*) and sentences (e.g., *I don't think so*), which over 95% of the participants got correct on the pretest, were excluded to minimize the ceiling effect. The participants' proficiency levels (A, B, C, and D) were set as their overall dictation score (x) ≥ 0.8 for Level A, $0.7 \leq x < 0.8$ for Level B, $0.6 \leq x < 0.7$ for Level C, and $x < 0.6$ for Level D.

Factor related to participants' attitudes toward English and task - For the EXP participants, class type (S&T or H&B; see Table 2 for details) and their attitudes toward English and the intervention were examined by the participants' ratings of their (a) attitudes toward English, (b) perceived awareness levels of intonation and

unstressed elements in the pre-questionnaire, and (c) usefulness of the dictation activity in the post-questionnaire. All scores were defined as the distance between the left end of the scale and the point the participant marked divided by the total length of the scale (see Figure 1).

In order to examine the significance of difference, we performed a two-way analysis of variance (ANOVA) on the rating scores with the main factors of participants' English level and class type. The mean scores for H&B classes were higher than those for S&T classes for all questions related to attitudes and awareness, and all differences between the two class types were statistically significant except for that for the awareness level of unstressed elements and for how they enjoyed dictation (see Table 5). The results suggest that the H&B-class participants, as a group, started with more positive attitudes toward English and with higher awareness levels, and they acknowledged the activity more highly. Based on these findings, there was reasonable evidence to suggest class type was a factor that was related to the participants' attitudes toward English and task.

Table 5

ANOVAs on Ratings of Questions in Pre-Post Questionnaires with Factors of Class-type and English-level

	Class type[a]	English level[a]	multiple comparisons[b]
Attitudes toward English in pre-questionnaire (n = 565)			
I like English	****	****	A > B > C ≥ D
English is important	***	***	A ≥ B > C ≥ D
I'm not good at English	*	****	A < B ≤ C ≤ D; B < D
Self-reported awareness in pre-questionnaire (n = 565)			
of Intonation	****	****	A > B ≥ C ≥ D; B > D
of Unstressed elements	p = .42	**	A ≥ B ≥ C ≥ D; A > D
Reflections on dictation in post-questionnaire (n = 348)			
Enjoyed?	p = .79	****	A ≥ B > C ≥ D
Recommend?	***	*	A ≥ D ≥ B ≥ C; A > B; A > C
Perceived usefulness of dictation in post-questionnaire (n = 348)			
Useful for listening	**	*	A ≥ B ≥ D ≥ C; A > C
Useful for speaking	****	p = .41	

Note. [a] the symbols in the column mean that the difference was significant by ANOVA tests: * is for $p < .05$, ** for $p < .01$, *** for $p < .001$, and **** for $p < .0001$;
[b] > and < in this column mean that the difference was significant, and ≥ and ≤ mean that the difference was non-significant, by Tukey's *post hoc* tests.

Participants' UF progress and factors: Task, English-level, and Class-type - Table 6 shows the summary of UF dictation accuracy rates in the pretest and posttest by task, participants' English level, and class type. A three-way ANOVA was performed to analyze the effect of UF progress with Task, English-level, and Class-type. In this analysis, the data of Level A participants were excluded to minimize ceiling effects. For all analyses, the alpha level of significance was set to .05.

Table 6

Descriptive Statistics of UF Dictation Accuracy Rates in Pre-Post Tests by Task, Class Type and English Level

Control group						Experimental group					
English level	n	$M1$[a]	$(SD1)$	$M2$[b]	$(SD2)$	English level -task	n	$M1$	$(SD1)$	$M2$	$(SD2)$
S&T classes											
A	7	.74	(.05)	.74	(.05)	A-Dg	16	.78	(.07)	.81	.(08)
						A-DgS	15	.79	(.07)	.82	(.08)
B	11	.67	(.07)	.64	(.10)	B-Dg	19	.63	(.05)	.68	(.07)
						B-DgS	19	.65	(.07)	.67	(.10)
C	20	.50	(.07)	.56	(.10)	C-Dg	11	.56	(.05)	.62	(.09)
						C-DgS	16	.52	(.07)	.59	(.11)
D	42	.33	(.10)	.38	(.16)	D-Dg	10	.35	(.07)	.46	(.13)
						D-DgS	11	.36	(.10)	.46	(.16)

Table 6 … continued

Control group						Experimental group					
English level	*n*	*M1*[a]	*(SD1)*	*M2*[b]	*(SD2)*	English level -task	*n*	*M1*	*(SD1)*	*M2*	*(SD2)*
H&B classes											
A	28	.80	(.08)	.87	(.06)	A-Dg	35	.78	(.07)	.81	(.10)
						A-DgS	43	.81	(.08)	.84	(.09)
B	43	.64	(.06)	.67	(.09)	B-Dg	33	.64	(.06)	.72	(.09)
						B-DgS	41	.62	(.05)	.69	(.10)
C	38	.52	(.06)	.61	(.09)	C-Dg	23	.53	(.07)	.61	(.10)
						C-DgS	19	.52	(.07)	.62	(.12)
D	29	.36	(.09)	.47	(.12)	D-Dg	21	.38	(.09)	.51	(.09)
						D-DgS	28	.35	(.12)	.46	(.12)

Note. [a]*M1* and *SD1* are the mean and standard deviation of UF dictation accuracy rates in the pretest;
[b]*M2* and *SD2* are the mean and standard deviation of UF dictation accuracy rates in the posttest; and
[c]task consists of two groups of participants: Dg did dictation and group work and DgS did dictation, group work, and shadowing.

The statistics revealed that Task [$F(2, 415) = 5.98$, $p = .003$], Class-type [$F(1, 415) = 10.15$, $p = .002$], and English-level [$F(2, 415) = 14.16$, $p < .001$] were significant, but interactions between those factors were non-significant. For Task, both experimental groups, Dg and DgS, showed significantly more UF progress than the control group. For the two experimental groups, the mean UF progress for the Dg group was slightly higher than that for the DgS group, but the difference was non-significant, $p = .742$ by a Tukey's *post hoc* test. For Class-type, the mean UF progress for the H&B group was higher than that for the S&T group.

For English-level, the mean UF progress of Level D participants was the largest and that of Level B participants was significantly lower than those of the lower level participants (B < C, $p = .024$; B < D, $p < .001$; Tukey's *post hoc* tests), but the difference between Levels C and D was non-significant ($p = .349$). Considering that the index (i.e., UF progress) used to measure the amount of UF progress, these results are comprehensible; low dictation accuracy scores in the pretest may simply mean that there was more room for improvement. While the Level D participants made the most UF progress, their post-UF accuracy rate was lower than Level C group's pre-UF accuracy rate. Therefore, we investigated which UF items participants, by proficiency level and task, could make the most progress on.

3.2 Progress of UF items related to participants' English level and task

UF item ordering in terms of difficulty - First, Pearson's correlation coefficients were calculated on dictation accuracy rates for each UF item (from UF01 to UF33) in the pretest averaged across all the participants at each English level (from A to D) (see Table 7). It was found that there were significant, high correlations between each English level's mean pre-accuracy rate scores (the minimum $r > .74$ for between the scores of the participants who were at Level A and those at Level D), suggesting that there was UF item ordering in terms of difficulty for this sample population regardless of English level.

Table 7

Pearson's Correlation Coefficients: Pre-Dictation Accuracy Rates for 33 UF Items by English Level

English Level	A	B	C	D
A	1.0000	.9639	.8982	.7422
B	-	1.0000	.9700	.8491
C	-	-	1.0000	.9229
D	-	-	-	1.0000

Note. All correlation coefficients were statistically significant ($p < .0001$).

UF items improved by task - To examine whether a difference in improved UF items between the CTL and EXP groups related to the participants' English levels existed, we calculated the mean dictation accuracy rate for each UF item, in both the pretest and posttest, averaged across task group (CTL or EXP) and English-level group

(A, B, C, or D). For articles, category matching (instead of exact matching) was employed when either articles (*a* or *the*) could be used in the context because Japanese learners tend to have difficulty in hearing the difference because the articles in this study are in weak forms and the sound [ð] in *the* does not exist in the Japanese language; thus, 0.5 points instead of 0 points were given if a participant wrote *a* for *the* and vice versa.

Each UF item's accuracy rate progress (UF item progress) was calculated for each participant as the UF item dictation accuracy rate in the posttest minus that in the pretest. Figure 2 shows the amount of each UF item progress by English level and task; the UF items are presented from the lowest to highest of each UF item's mean dictation accuracy rate in the pretest averaged across the four English levels. In order to examine whether the differences between the CTL and EXP groups were significant, we performed a non-parametric test (Kruskal-Wallis) on each participant's UF item progress with task (CTL or EXP). The non-parametric test was selected because each UF item's dictation point awarded for a participant was either 1 (= correctly recognized) or 0 (= missed or error) or 0.5 (for articles). The statistically significant difference ($p < .05$) was indicated with '(*)' in Figure 2.

The Level A participants in the EXP group did not show any substantial improvements, but participants who were at Levels B, C, and D in the EXP group showed the following statistically significant progress, compared to those in the CTL group: articles (*a*, *the*), preposition (*to*), and plural (*–s*). In particular, Levels B and D revealed the most significant progress. The UF item progress for each Level created a clear division (see Figure 2).

UF item	mean AR1[a]	Level D			Level C			Level B			Level A		
		AR1[b]	CTL	EXP	AR1	CTL	EXP	AR1	CTL	EXP	AR1	CTL	EXP
UF26	.06	.01			.02			.04			.19		
UF30	.12	.04			.04	+		.11		+ (*)	.29	++	+
UF08	.14	.01			.02			.10		+ (*)	.42	++	++
UF25	.15	.06			.09			.14			.32	+	
UF07	.16	.01			.03			.11		+ (*)	.47	++	+
UF06	.18	.01			.04			.13		+	.55	++	+
UF17	.19	.04			.09	+		.17	+		.46		
UF16	.23	.01			.06	+	+	.29		+	.56	+	
UF27	.38	.13			.27	+		.48			.65		
UF23	.50	.14			.35		+	.63			**.88**		
UF09	.51	.15	++	++	.38	++	+	.61	+	+?	**.88**		
UF03	.52	.20			.37	+	+	.64	+	+	**.87**	+	
UF28	.52	.32			.48			.56			.71		
UF12	.55	.21		++	.47	+	++	.63	+	+	**.87**		
UF24	.62	.36			.58		+	.73			.79		
UF33	.63	.38	+	++	.48	++	+	.74			**.90**	+	
UF05	.64	.33	+	++	.62		+	.72			**.87**		
UF13	.65	.28	++	++	.63	+	+	.78			**.92**		
UF21	.66	.26	+?	+	.64	+	+	**.81**			**.95**		
UF10	.67	.32	+	++	.69			.77		+	**.91**		
UF02	.70	.36	+	++	.69	+	+	**.80**			**.95**		
UF32	.70	.41	+	+	.65			**.82**			**.94**		
UF22	.76	.39		++	.75	+	+	**.89**			**.99**		
UF14	.78	.60		++ (*)	.77	+		**.83**		+	**.94**		
UF11	**.80**	.46	+	+	**.84**		+	**.91**			**.99**		
UF29	**.82**	.62	+	++	**.82**		+ (*)	**.89**			**.96**		
UF31	**.82**	.58		+	**.81**			**.95**			**.95**		
UF20	**.83**	.62		+++ (*)	**.84**		+	**.89**			**.96**		
UF19	**.89**	.68		+++ (*)	**.93**			**.96**			**.99**		
UF18	**.92**	**.87**			**.89**			**.96**			**.98**		
UF01	**.94**	**.93**			**.94**			**.93**			**.96**		
UF15	**.95**	**.88**			**.95**			**.98**			**1**		
UF04	**.98**	**.92**			**.99**			**.99**			**1**		

Note. [a]mean-AR1 is mean dictation accuracy rate of each UF item averaged across the four English levels; [b]AR1 is the mean value for dictation accuracy rate of each UF item in the pretest; the figures in bold are higher than 0.8; + shows the amount of UF-item progress; +? = 9.756%, 10% <= + < 20%, 20% <= ++ < 30%, and 30% <= +++; and (*) shows that the difference between the groups (CTL and EXP) was significant (p < .05; Kruskal-Wallis tests).

Figure 2. Amount of dictation accuracy progress for UF items by English level and task, sorted by each UF item's mean dictation-accuracy rate in the pretest across four English levels (mean-AR1)

In the case of articles, most participants were able to make progress on the posttest, but the context in which the articles were in varied in difficulty, which was related to their proficiency levels. As is seen from Figure 2, the Level B participants in the EXP group made improvement on UF30: *Why don't we rent **a** DVD?* and UF07: *I believe he's **the** last man to do such a thing* significantly more than the CTL group, but this was not the case for the Level D participants. These UF items were too difficult for the Level D participants to show any improvement; their accuracy rates in the pretest were only 4% for UF30 and 1% for UF07. However, they made significant progress on UF20: *Yes, **the** new class monitor will be elected soon* and UF19: *Who is **the** new class monitor?* their accuracy rates for which were reasonably high (62% for UF20 and 68% for UF19).

3.3 Self-reported awareness levels in pre-post questionnaires

Relationship between self-reported awareness levels and dictation performance - We examined whether higher awareness scores were related to higher dictation performance. A Pearson's correlation coefficient (r) was separately calculated between each participant's overall dictation scores in the pretest and (a) self-reported awareness scores of intonation and (b) those of unstressed elements in the pre-questionnaire. The self-reported awareness score was calculated as the distance between the left end of the scale and the point the participant marked divided by the total length of the scale.

The self-reported awareness of intonation ($r = .2745$, $p < .0001$) was significant but had a weak correlation, and the self-reported awareness of unstressed elements ($r = .1582$. $p = .0002$) was also significant, but r was very small; therefore, we decided to use levels instead of raw scores for this variable of awareness of unstressed elements (AUF). The three levels of AUF (AUF levels) were defined as follows: Low (AUF < 0.5), Middle (AUF = 0.5), and High (AUF > 0.5). The results showed that, for the overall dictation scores in the pretest, both main factors of Class-type [$F(1, 560) = 7.04$, $p = .008$] and self-reported awareness of unstressed elements in the pre-questionnaire (AUF1-level) [$F(2, 560) = 4.57$, $p = .011$] were significant. For the UF dictation accuracy scores in the pretest, the main factors of Class-type [$F(1, 560) = 5.25$, $p = .022$] and AUF1-level [$F(2, 560) = 3.02$, $p = .0496$] were also significant. For both analyses, the interactions between Class-type and AUF1-level were non-significant. Participants who had higher dictation performance in the pretest were more likely to report higher levels of awareness of UF.

Change in participants' self-reported awareness scores in pre-post questionnaires - Because this study focused on unstressed elements, we only discuss the self-reported awareness of UF hereafter. Table 8 shows the summary of the self-reported awareness levels of UF in the pre-post questionnaires by class type and participants' English level.

Table 8

Rating Scores for Self-Reported Awareness of Unstressed Elements in Pre-Post Questionnaires

English level	Control group						Experimental group					
	n	*M1*[a]	*(SD1)*	*M2*[b]	*(SD2)*	prog-ress[c]	*n*	*M1*	*(SD1)*	*M2*	*(SD2)*	prog-ress
S&T classes												
A	7	.43	(.27)	.54	(.17)	0.12	30	.48	(.22)	.48	(.23)	0
B	9	.26	(.26)	.48	(.19)	0.23	37	.36	(.24)	.43	(.19)	0.07
C	20	.29	(.25)	.48	(.20)	0.19	26	.34	(.22)	.48	(.23)	0.14
D	41	.33	(.29)	.39	(.24)	0.06	21	.36	(.22)	.44	(.28)	0.08
H&B classes												
A	28	.45	(.19)	.54	(.31)	0.09	74	.42	(.24)	.60	(.21)	0.18
B	41	.42	(.27)	.58	(.23)	0.16	73	.39	(.23)	.53	(.24)	0.15
C	37	.38	(.27)	.52	(.23)	0.14	41	.40	(.26)	.43	(.24)	0.04
D	28	.37	(.20)	.47	(.13)	0.10	48	.32	(.25)	.50	(.23)	0.18

Note. [a]*M1* and *(SD1)* are the mean and standard deviation of scores of self-reported awareness of unstressed elements (AUF) in the pre-questionnaire; [b]*M2* and *(SD2)* are the mean and standard deviation of AUF scores in the post-questionnaire; and progress is *M2* minus *M1*.

To examine what factor(s) contributed to raising some participants' AUF scores, the change (AUF change) in the participants' AUF scores in the post-questionnaire minus that in the pre-questionnaire were used. Then, we performed a four-way ANOVA on each participant's AUF change with main factors of task, class type and participants' English levels, and the level of pre-questionnaire AUF-score (AUF1). Two levels of AUF1 (AUF1-level-L) were defined as follows: Low (AUF1 < 0.5), and non-Low (AUF1 ≥ 0.5). The variable of AUF1-level-L was reduced from 3 (Low, Middle, and High) used in the previous analysis to 2 [Low and non-Low (Middle and High pooled)] because (a) there were 5 cells with only 1 occurrence; and (b) the AUF1 means for the Middle and High groups were not statistically different. The data with AUF1 ≥ 0.8 were excluded to minimize the ceiling effects.

The main factors of Class-type [$F(1, 508) = 1.17$, $p = .280$], English-level [$F(3, 508) = 0.29$, $p = .832$], and Task [$F(1, 508) = 0.06$, $p = .807$] and all interactions were non-significant. The only significant factor was AUF1-level-L (whether AUF1 was low or not), $F(1, 508) = 100.53$, $p < .0001$; more participants who answered "I do not pay much attention to the UF while listening to English" in the pre-questionnaire came to acknowledge the need of paying more attention to UF in the post-questionnaire compared to those who reported higher awareness levels from the beginning.

Because the level of AUF1 was the only decisive factor to account for the change in the self-reported awareness level in this analysis, we also performed a two-way ANOVA on scores of AUF change from the participants whose AUF1 = 0.5 in the EXP group. The main factors used for the analysis were Class-type and English-level. The results showed that the mean AUF change for the H&B group was higher ($M = .06$, $SD = .23$) than that for the S&T group ($M = -.03$, $SD = .23$), but the difference was non-significant ($p = .149$). The English-level difference was also non-significant ($p = .515$). However, the interaction between Class-type and English-level was found to be significant, $F(3, 130) = 3.56$, $p = .016$. The higher-level participants in H&B classes showed substantial change in the self-reported awareness level; the mean AUF change score was .15 for Level A, .10 for level B, but .02 for Level D, and -.10 for Level C. In contrast, the overall pattern for the AUF change by English level found in S&T classes was significantly different from that in H&B; only the Level C participants in S&T classes showed positive change in the self-reported awareness level (mean AUF change = .08), but the mean scores were all negative for the other level groups.

Effect of self-reported awareness scores of UF on task performance - To examine whether participants with high awareness in the pre-questionnaire showed better task performance, we carried out a two-way ANOVA on UF progress of the lower-level (Levels C and D pooled) participants in the H&B group with the main factors of task and initial awareness level. A main factor of the initial level of self-reported awareness (AUF1-level-H) was defined as follows: High (AUF1 > 0.5) and non-High (AUF1 ≤ 0.5). We used only data of lower-level participants in H&B classes because participants' English level and class type were found to be significant factors accounting for the difference of UF dictation progress. The lower-level participants were selected because they had more room for improvement. The data of the participants who were at Levels C and D were pooled because the mean UF progress scores for both groups were not significantly different (by Tukey's *post hoc* test, $p = .113$). H&B classes were selected because a greater difference between the CTL and EXP groups was expected for H&B classes than for S&T classes. In-class dictation exercises were not employed in H&B classes but were used as part of regular S&T classes.

For both CTL and EXP groups, the mean UF in the pretest was higher for the High group than the non-High group. As mentioned, the participants starting from lower UF scores tended to show more UF dictation accuracy improvement, so it was expected that the mean UF progress for the High group would be lower than that for the non-High group. The result of the CTL group was just as predicted; UF progress was higher for the non-High group than for the High group. However, the EXP participants, who received the awareness-raising intervention, showed greater progress for the High group than the non-High group (see Figure 3). This suggests that the participants who started with higher self-reported awareness levels for UF benefited from the task greatly in terms of UF dictation performance.

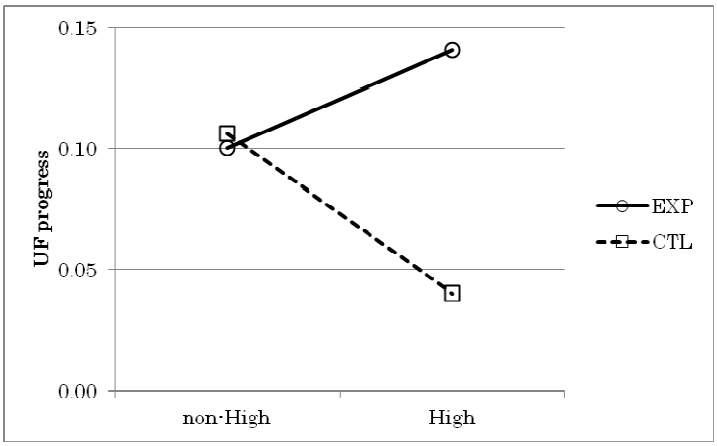

Figure 3. UF progress of non-High and High groups

3.4 Participants' reflections on the activities in the post-questionnaire

The participants in the EXP group were asked in the post-questionnaire (Appendix B) to mark on the scales (see Figure 1 for a sample) for the questions concerning the activities they did. The EXP participants' rating scores were calculated as the distance between the left end of the scale and the point the participants marked divided by the total length of the scale, thus it ranged from 0 (meaning *I don't think so at all*) to 1 (meaning *I think so very much*). Table 9 shows the means and standard deviations of the participants' self-reported enjoyment and usefulness of the activities. According to the ratings, the participants seemed to enjoy all three components of the intervention task and considered them useful for both listening and speaking. We separately performed a two-way ANOVA on each participant's rating score for all questions with the main factors of Class-type and English-level (see Table 10 for summary of results).

Table 9

Means and SDs of Rating Scores in Post-Questionnaire

	n	*M*	*SD*
Did you enjoy the activity?			
Q1-1: dictation	349	.70	.24
Q1-2: group work	349	.75	.23
Q1-3: shadowing	188	.68	.23
Was the activity useful for listening?			
Q3-1: dictation	350	.88	.16
Q3-2: shadowing	189	.86	.18
Was the activity useful for speaking?			
Q3-3: dictation	350	.81	.19
Q3-4: shadowing	189	.83	.21

Table 10

ANOVAs on Participants' Reflections of Activities in Post-Questionnaire by Class Type and English Level

Questions	Class type	English Level
1-1: Enjoyed dictation?	[S&T < H&B]	**[A ≥ B > D ≥ C]
1-2: Enjoyed group work?	**[S&T < H&B]	[B ≥ A ≥ D ≥ C]
1-3: Enjoyed shadowing?	(p = .051)[S&T <= H&B]	[D ≥ A >= B >= C]
2-1: Was dictation easy?	[S&T < H&B]	**[A ≥ B ≥ C > D]
2-2: Was shadowing easy?	[H&B < S&T]	[A ≥ C ≥ B ≥ D; A > D]
3-1: Useful? dictation for listening	**[S&T < H&B]	*[A ≥ B ≥ D ≥ C; A > C]
3-2: Useful? dictation for speaking	**[S&T < H&B]	[D ≥ A ≥ B ≥ C]
3-3: Useful? shadowing for listening	** [S&T < H&B]	[A ≥ D ≥ B ≥ C; A > C]
3-4: Useful? shadowing for speaking	** [S&T < H&B]	[D ≥ A ≥ B ≥ C; D, A > C]

Table 10 … continued

Questions	Class type	English Level
4-1-Dg: Recommend dictation and group work? <Dg group only>	[S&T < H&B]	[A ≥ D ≥ B ≥ C]
4-1: Recommend dictation? <DgS group only>	** [S&T < H&B]	[A ≥ D ≥ B ≥ C]
4-2: Recommend group work? <DgS group only>	** [S&T < H&B]	[D ≥ A ≥ C ≥ B]
4-3: Recommend shadowing?	** [S&T < H&B]	[A ≥ B ≥ D ≥ C; A > C]

Note. * and ** show that the difference was significant by ANOVA tests ($p < .05$ for *; $p < .01$ for **); and ≥ in the English-level column means that the difference was non-significant ($p > .05$) while > means that the difference was significant.

General comparisons - The results showed that the scores of H&B classes were higher than those of S&T classes for all questions except Q2-2: *How easy was shadowing?*. Level A participants' scores were the highest or second highest among all four groups of participants in all questions, and those of Level C participants were the lowest for 10 questions out of 12. For Level D participants, the scores were the highest for the following questions: (a) *Did you enjoy shadowing?* (Q1-3); (b) *Do you think the activities of dictation and shadowing were useful for speaking?* (Q3-2 & Q3-4), and (c) *Would you recommend group work?* (Q4-2). The Level D participants' responses indicate that their favorite activities were shadowing, speaking, and group work, which we assume were relatively easier for them than the dictation.

Comparisons between dictation and group work - Table 11 shows the summary of the means (*M*s) and standard deviations (*SD*s) of the EXP participants' rating scores for the questions about how much they enjoyed the activities (*Enjoyed* scores), Q1-1 for dictation, and Q1-2 for group work. The mean of Level A participants' Enjoyed scores was higher ($M = .80$) for dictation than that for group work ($M = .75$), though it was not statistically significant by t-test ($p = .101$). However, the mean scores of the participants who were at Levels B, C, and D were higher for group work than those for dictation, and according to 2-tailed t-tests, the differences were statistically significant for the groups of levels C and D ($p < .001$) but did not reach the significant level of .05 for the Level B group ($p = .075$).

Table 11

Statistics for Enjoyed Scores for Dictation (Q1-1) and for Group Work (Q1-2) by English Level and Difference

English level	*n*	Dictation (Q1-1) *M*	(*SD*)	Group work (Q1-2) *M*	(*SD*)	*p*
A	102	.80	(.21)	.75	(.25)	.101
B	110	.73	(.23)	.77	(.22)	.075
C	67	.60	(.19)	.71	(.22)	**.0006**
D	69	.63	(.28)	.75	(.21)	**.0003**

Comparisons between dictation and shadowing - Table 12 shows the summary of the means and standard deviations of the EXP-DgS participants' rating scores for the questions about their perceived easiness of the activity (*Easy* scores), Q2-1 for dictation and Q2-2 for shadowing. For the participants of all English levels, as was expected, the Easy scores for shadowing were higher than those for dictation (see Table 12). We then carried out 2-tailed t-tests. The results revealed that the difference was significant for the participants who were at Levels A, B, and D ($p < .05$) but non-significant for Level C participants ($p = .089$).

Table 12

Statistics for Easy Scores for Dictation (Q2-1) and for Shadowing (Q2-2) by English Level and Difference

English level	*n*	Dictation (Q2-1) *M*	(*SD*)	Shadowing (Q2-2) *M*	(*SD*)	*p*
A	57	.46	(.18)	.53	(.17)	**.011**
B	60	.37	(.19)	.46	(.21)	**.005**
C	34	.40	(.18)	.47	(.17)	.089
D	38	.26	(.26)	.41	(.24)	**.0001**

Table 13 shows the summary of the means and standard deviations of the EXP-DgS participants' rating scores for the questions about how much they enjoyed (Enjoyed scores) and how highly they would recommend (*Recommend* scores) the dictation and shadowing activities. We did 2-tailed t-tests, and the significance of difference was also included in Table 13. Although the differences were not significant except the Recommend scores for the Level B group, we found a unique pattern for Level D participants in this sample. Level D participants tended to report that they enjoyed shadowing (M = .71) more than dictation (M = .62) but gave similar scores to dictation (M = .75) and to shadowing (M = .73) when they were asked how highly they would recommend these activities. The other groups of participants (Levels A, B, and C), however, tended to answer that they enjoyed dictation more than shadowing but would recommend shadowing more highly.

Table 13

Statistics for Enjoyed Scores for Dictation (Q1-1) and for Shadowing (Q1-3), and Recommend Scores for Dictation (Q4-1) and for Shadowing (Q4-3) by English Level, and Difference

English level	*n*	Enjoyed? Dictation (Q1-1) *M*	(*SD*)	Shadowing (Q1-3) *M*	(*SD*)	*p*	Recommend? Dictation (Q4-1) *M*	(*SD*)	Shadowing (Q4-3) *M*	(*SD*)	*p*
A	56	.77	(.21)	.71	(.24)	.090	.77	(.20)	.80	(.19)	.582
B	60	.70	(.26)	.68	(.22)	.530	.70	(.22)	.76	(.18)	**.009**
C	34	.60	(.19)	.59	(.19)	.971	.68	(.21)	.69	(.26)	.863
D	38	.62	(.29)	.71	(.23)	.109	.75	(.19)	.73	(.20)	.404

4. Discussion

Our first research question was: To what extent did the awareness-building and collaborative intervention influence participants' dictation performance of unstressed elements? For UF progress, the activities were effective, but the difference between the participants of the two EXP groups (i.e., Dg and DgS) was not statistically significant. We expected that the DgS participants would outperform the Dg participants in dictation performance progress because the DgS participants did an extra activity of shadowing as well as the same activities done by the Dg group. The reasons may be (a) the amount and period of shadowing might not have been sufficient to produce positive results; or (b) many of the Dg participants might have been doing subvocalization during their dictation activities (e.g., Postovsky, 1974), which might have created similar effects to those created by real shadowing.

As was stated in the third research question, we also examined the factors of the participants' attitudes toward English and task (by using a variable of Class-type, categorized as More/Less positive attitudes), and of their proficiency levels (based on the pretest), related to the participants' post-performance of unstressed elements (UF). The attitudinal factor was significant for UF progress; more progress was seen in the H&B group, which was categorized as *having more positive attitudes*, than the S&T group. More positive attitudes toward English and task may have led to more progress of UF dictation performance (e.g., Kormos & Dörnyei, 2004). The proficiency factor was also significant for UF progress. The participants in the lower-level group made the most progress in terms of UF accuracy rate on the posttest; however, those low-level participants' mean UF

accuracy rate did not reach the upper-level participants' mean pretest UF accuracy rate.

Improved UF components were examined related to the participants' proficiency levels. High correlations were found between the UF item ordering in terms of difficulty at all levels (see Table 7). In other words, which UF item was more difficult in dictation did not so much depend on the proficiency level. This is comprehensible because learners experience common problems when completing dictation tasks, including (a) unknown/unfamiliar syntax, expressions and words; (b) unstressed sound; (c) sound change by coarticulation (e.g., blending); and (d) the length of the sentence/chunk used in dictation (requiring larger capacity of working memory) (e.g., Field, 2003; Roach, 2009; Underhill, 2005). Each group moved up the hierarchy consistently indicating that if selected unstressed grammatical elements were placed in appropriate context, as in this study, there would be a possibility of using this kind of dictation task for a level placement test.

It was found that the Level A participants in the EXP group and CTL group showed similar improvements. One reason may be that the Level A participants' awareness levels may have been high from the beginning and thus were not influenced by the awareness-raising intervention. Another reason may be the ceiling effect; the dictation sentences used in the pretest and posttest may have been too easy to indicate progress. Although the Level A participants and the other level participants received the same intervention, the other level participants were able to build on their prior knowledge and base skills to make progress. In particular, the lower-level groups made the most progress, especially in listening to unstressed elements in word-final positions and articles; most of the participants could not hear articles, although they were thought to have grammatical knowledge of articles from their previous compulsory English classes. However, after the awareness-raising intervention, they could hear the articles if those articles were positioned in simple sentences with words that were familiar to them.

Most of the lower-level group participants did not make progress for more difficult UF items for which the higher-level participants did. One possible explanation could be that those participants couldn't spare enough attention for the unstressed elements because they were overloaded with focusing on key words, even if they had acknowledged the importance of paying attention to unstressed elements. Thus, it could be assumed that the lower-level participants were able to gain awareness of unstressed elements appropriate to their level at that time and could make further progress.

Awareness-raising tasks do have limitations; we cannot expect the learners' vocabulary size or grammatical/phonetic knowledge, let alone the working memory capacity, to be expanded only by an awareness-raising task in a short period of time. In a study done by Williams and Evans (1998) with 33 ESL university students in the U.S. on focus-on-form of one difficult and one easy grammatical feature, the participants' readiness seemed to influence the results. In other words, the participants who had internalized that knowledge partially as evident by their pretest scores might have benefitted from instruction. The results of the study showed that the participants with low pretest scores were only able to make progress on the easy grammatical feature, and a few participants in the control group who most likely had partial prior mastery of the target grammatical features were able to make progress even on the difficult grammatical feature without instruction. However, our results showed that the awareness-raising tasks might have helped the participants, especially those who were lower-level, notice the features by giving them experiences to anticipate where an article or preposition in a certain context should be or of paying attention to word-final positions. Thus, these results show that it is important for educators to create awareness-raising tasks with students' readiness in mind.

Our second research question was: Did the intervention tasks have an influence on participants' self-reported strategic use of paying attention to unstressed elements? The factors of both attitude and proficiency level were examined together with task (see third research question). We used a variable of self-reported awareness scores of unstressed elements in the spoken input (AUF) in both pre-post questionnaires, ranging from 0 (not at all) to 1 (very much so). AUF was a score related to how often/much a learner would use this learning strategy of paying attention to unstressed elements, (which was categorized as a "metacognitive strategy" of selective attention in

Zafar & Meenakshi, 2012, p. 642).

Both the overall and UF dictation accuracy rates in the pretest and their self-reported awareness scores in the pre-questionnaire were positively related. In other words, the participants who demonstrated higher dictation performance tended to report higher levels of AUF. In addition, it was found that with the same intervention task, the participants with the higher AUF scores in the pre-questionnaire made more progress than those with the lower scores. These results seem to suggest the importance of awareness-raising, such as the strategic use of selective attention in this study in learning language.

Factors of Task, Class-type, and English-level did not have any significant effect on changes in AUF scores. However, there was a significant interaction between Class-type and English-level among the participants with pre-AUF = 0.5. The higher-level participants in H&G classes had increased post-AUF scores compared to lower-level ones, but this pattern was not seen for S&T classes. In general, H&G class participants reported more positive attitudes toward English and task than the S&T class ones. This might suggest that attitudinal factors could have influenced the change in awareness scores.

With the awareness-raising intervention, the participants did make progress in dictation performance, especially in such UF items as articles and word-final morphological items, but the intervention failed to raise the self-reported awareness score (AUF) contrary to our expectation. One reason might be that some participants who did not report increased strategic use of paying attention to unstressed elements might have done so unconsciously. Another reason might be that some participants who started to pay attention to unstressed elements may not have necessarily linked what they did with a learning strategy.

Finally, we examined how the other components of the intervention task were perceived by the participants in relation to dictation. Level A is the only group that enjoyed diction more than group work. This could be because the sentences used in the dictation activities were not difficult for them so that they could do the task by themselves. The lower-level participants thought group work was more enjoyable than dictation. This could be because group work made them feel *safe* and *relaxed*, as some of the participants reported in the post-questionnaire, because they could see that others in their group had conflicting answers during the discussions, which is not possible for a teacher-centered classroom approach. The positive results in this study are consistent with Morris and Tremblay's (2002) finding of small group discussion. In addition, Slavich and Zimbardo's (2012) findings suggest that collaborative learning is effective when students challenge complex issues in groups, and it is best accomplished when at least one member of the group is more knowledgeable than the others. However, some participants reported and instructors observed that active and positive group work depended on the interaction between the members in the groups; for example, some participants did not want to participate in the discussions, so they would simply copy the most advanced student's answers for the final versions of the cloze dictation answers.

Shadowing was perceived as easier than diction for all groups. Only the lower-level participants thought that shadowing was more enjoyable than dictation, maybe because it was thought to be non-threatening and accomplishable (i.e., they could succeed in completing the task). According to Sharwood Smith and Truscott (2014), learners' attitudes toward inputs could affect what they decide to do or not to do. For example, they refer to the negative effects of anxiety and the possibility of how emotion could influence language learning. This could imply that for lower-level learners, in particular, a task that is challenging but can be accomplished is an important factor, and thus educators should take the learners' language proficiency into consideration in designing tasks. This means that if accomplishable activities such as group work and shadowing are included in a task that may be slightly too challenging to some lower-level learners, it might become more satisfying and enjoyable for them, which may lead to more positive formative and summative outcomes.

The participants who were at Levels A, B, and C thought that dictation was more enjoyable than shadowing, but they recommended shadowing more than dictation; the complete opposite was true for level D. In other words, many learners claimed that enjoying a task is not the same as making progress and do not always

recommend the task to be repeated for peers, as found in Crouch and Mazur's (2001) study on a classroom approach for university students studying physics.

5. Conclusions

This large-scale study demonstrated that the cloze dictation activities focusing on unstressed grammatical elements and group work were determined to be effective in helping the EFL learners improve their listening ability. However, the effectiveness of shadowing was inconclusive. The participants' class type used as an attitudinal factor and their proficiency levels, which were based on their pretest scores, were also found to influence their post-dictation performance. Participants who were interested in English and the intervention tasks were more likely to make progress, and those who had low pre-dictation scores made the most progress but were unable to make the same progress on UF items, which were considered difficult, as the participants in the upper-level groups.

After the pre-dictation test, a hierarchy of UF items was discovered. This means that all of the participants could master a few of the UF items, which were labeled *easy*, but only the participants who had high pre-dictation scores could master the more *difficult* ones. It was interesting to note that a clear UF item ordering in terms of difficulty or hierarchy resulted. After the post-dictation test, participants in the experimental group made more statistically significant progress on certain UF items compared to the control group. However, even though all of the participants in the experimental group were given the same cloze-dictation activities, the UF items that they improved on varied depending on their pre-dictation scores.

We also examined dictation performance in relation to participants' strategic use of paying attention to unstressed elements, by analyzing their awareness rating scores (referred to as AUF) in the pre- and post-questionnaires. It was found that AUF in the pre-questionnaire was related to their dictation scores in the pretest. Also, the participants who started with higher AUF scores demonstrated higher task performance, suggesting they benefited from the task more than the participants with lower pre-AUF scores. The results point to the importance of raising AUF scores, but the effect of the intervention in addition to the participants' proficiency level and class type were not significantly related to increased, self-reported AUF scores with the exception of the pre-AUF scores. However, after analyzing the data from participants who marked their AUF scores at the middle point on the rating scale, it was found that participants with more knowledge of unstressed elements and in classes with motivated peers tended to report an increased AUF.

There are some limitations to this study. Although a convenience sample was employed, the threats to external validity were minimized by the large number of participants with different majors from four institutions of higher education, facilitating generalizability of results (e.g., Mertens, 2015). To limit the threats to reliability of the participants' self-reported awareness scores in the questionnaires we used high and low levels from the mid-point on the scale instead of raw scores. However, future studies are needed to determine the reliability of the participants' other self-reported scores. In addition, the gender effects were not addressed because even though the number of females was larger than the number of males, both class types had similar male-female proportions. Also, while it is speculated that future studies involving high-level students might show that they employ self-regulated strategic use of paying attention to unstressed elements, we were unable to assess this factor and the influence of the tasks due to ceiling effects. Also, while the effects of shadowing on dictation performance could not be determined, it is speculated that the time and amount of the activity were insufficient and were overshadowed by the dictation activities. Many high-level participants, however, reported that the output activities were beneficial, which could be examined further in a future study. For dictation performance, prior knowledge or readiness to learn could influence progress as was demonstrated in this study, and delayed effects of the intervention might be evident over time. In addition, other attitudinal factors besides *the field of study* examined in this study might have influenced the change in both awareness and performance. Future studies could also investigate language learners' attitudinal factors and readiness to learn as well as delayed effects in relation to learning outcomes.

In summary, this study provides some important educational implications. First of all, the intervention in this study was received well by the participants, which is a motivating factor. Also, the intervention tasks could be implemented easily by educators, who use English as a first or foreign language, in a limited amount of time, for example, the first 15 to 20 minutes of class, during weekly or daily sessions with students of varying levels. Finally, if educators employed the pre-post dictation tasks, they could assess students' formative progress and summative outcomes.

6. References

Abadikhah, S., & Shahriyarpour, A. (2012). The role of output, input enhancement and collaborative output in the acquisition of English passive forms. *Journal of Language Teaching and Research*, *3*(4), 667-676. http://dx.doi.org/10.4304/jltr.3.4.667-676

Baars, B. J., & Gage, N. M. (2012). *Fundamentals of cognitive neuroscience: A beginner's guide*, Boston: Academic Press.

Chun, A., Day, R. R., Chenoweth, A., & Luppescu, S. (1982). Errors, interaction, and correction: A study of native-nonnative conversation. *TESOL Quarterly*, *16*(4), 537-547. http://dx.doi.org/10.2307/3586471

Crouch, A. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics*, *69*(9), 970-977. http://dx.doi.org/10.1119/1.1374249

DiCerbo, P. A., Anstrom, K. A., Baker, L. L., & Rivera, C. (2014). A review of the literature on teaching academic English to English language learners. *Review of Educational Research*, *84*(3), 446-482. http://dx.doi.org/10.3102/0034654314532695

Dörnyei, Z. (2006). Individual differences in second language acquisition. *AILA Review*, *19*, 42-68. http://dx.doi.org/10.1075/aila.19.05dor

Dörnyei, Z., & Kormos, J. (2000). The role of individual and social variables in oral task performance. *Language Teaching Research*, *4*(3), 275-300. http://dx.doi.org/10.1177/136216880000400305

Doughty, C., & Williams, J. (Eds.) (1998a). *Focus on form in classroom second language acquisition*. New York: Cambridge University Press.

Doughty, C., & Williams, J. (1998b). Pedagogical choices in focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 197-261). New York: Cambridge University Press.

Field, J. (2003). Promoting perception: Lexical segmentation in L2 listening. *English Language Teachers Journal*, *57*(4), 325-334. http://dx.doi.org/10.1093/elt/57.4.325

Guilloteaux, M. J., & Dörnyei, Z. (2008). Motivating language learners: A classroom-oriented investigation of the effects of motivational strategies on student motivation. *TESOL Quarterly*, *42*(1), 55-77. http://dx.doi.org/10.1002/j.1545-7249.2008.tb00207.x

Hammerly, H. (1987). The immersion approach: Litmus test of second language acquisition through classroom communication. *Modern Language Journal*, *71*(4), 395-401. http://dx.doi.org/10.1111/j.1540-4781.1987.tb00378.x

Jones, M. E., Antonenko, P. D., & Greenwood, C. M. (2012). The impact of collaborative and individualized student response system strategies on learner motivation, metacognition, and knowledge transfer. *Journal of Computer Assisted Learning*, *28*(5), 477-487. http://dx.doi.org/10.1111/j.1365-2729.2011.00470.x

Kormos, J., & Dörnyei, Z. (2004). The interaction of linguistic and motivational variables in second language task performance. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* [Online], *9*(2), 19 pp. Retrieved April 18, 2015, from http://zif.spz.tu-darmstadt.de/jg-09-2/beitrag/kormos2.htm

Lasagabaster, D., & Sierra, J. M. (2005). Error Correction: Students' versus Teachers' Perceptions. *Language Awareness*, *14*(2-3), 112-127. http://dx.doi.org/10.1080/09658410508668828

Mertens, D. M. (2015). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods* (4th ed.). Thousand Oaks, CA: Sage Publications, Inc.

Morris, L., & Tremblay, M. (2002). The impact of attending to unstressed words on the acquisition of written

grammatical morphology by French-speaking ESL students. *The Canadian Modern Language Review*, *58*(3), 364-385. http://dx.doi.org/10.3138/cmlr.58.3.364

Ofuka, E. (2015). *Awareness of unstressed functional elements: Error corrections in dictation exercises by Japanese EFL learners*, Unpublished manuscript**.**

Postovsky, V. (1974). Effects of delay in oral practice at the beginning of second language learning. *The Modern Language Journal*, *58*(5-6), 229-239. http://dx.doi.org/10.1111/j.1540-4781.1974.tb05104.x

Roach, P. (2009). *English phonetics and phonology: A practical course* (4th ed.). Cambridge: Cambridge University Press.

Roberts, M. A. (1995). Awareness and the efficacy of error correction. In R. Schmidt (Ed.) *Attention and awareness in foreign language learning* (Technical Report #9) (pp. 163-182). Honolulu, Hawai'i: University of Hawai'i, Second Language Teaching & Curriculum Center.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129-158. http://dx.doi.org/10.1093/applin/11.2.129

Sharwood Smith, M., & Truscott, J. (2014). Explaining input enhancement: A MOGUL perspective, *International Review of Applied Linguistics in Language Teaching*, *52*(3), 253-281.

Slavich, G. M., & Zimbardo, P. G. (2012). Transformational teaching: Theoretical underpinnings, basic principles, and core methods. *Educational Psychological Review*, *24*(4), 569–608. http://dx.doi.org/10.1007/s10648-012-9199-6

Slimani, A. (1992) Evaluation of classroom instruction. In J. Alderson & A. Beretta (Eds.), *Evaluating second language education* (pp. 197-221). Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139524575.009

Swan, M. (2005). *Practical English Usage* (3rd ed.). Oxford: Oxford University Press.

Underhill, A. (2005). *Sound foundations: Learning and teaching pronunciation*. Oxford: Macmillan.

Williams, J., & Evans, J. (1998). What kind of focus and on which forms? In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 139-155). New York: Cambridge University Press.

Zafar, S., & Meenakshi, K. (2012). Individual learner differences and second language acquisition: A review. *Journal of Language Teaching and Research*, *3*(4), 639-646. http://dx.doi.org/10.4304/jltr.3.4.639-646

Appendices:

Appendix A

Spring 2014

Name:_____ class:( ) student number:（ ）

1. ３か月以上の海外生活の経験はありますか。 YES / NO
   (Have you lived in another country for three months or longer?)

   もし YES の場合、どこにどのくらいの期間ですか。 (where_____ how long_____)
   (If YES, where did you live and for how long did you live there?)

2. 日常生活で英語を使うこと（家や寮に留学生がいる、家で英語を使う、など）がありますか。: YES / NO
   (Do you use English in your daily life? (For example, at home, at the dorm, with an exchange student)

   If YES, please explain:_____

3. Mark on the scale: (｜と ｜の間に印をつけてもＯＫ)

   No! Neutral Yes!

   ・英語が好き？ |---------|---------|---------|---------|
   (Do you like English？)

   ・英語に苦手意識がある？ |---------|---------|---------|---------|
   (Do you think you are not good at English？)

   ・英語はあなたの将来に重要？ |---------|---------|---------|---------|
   (Is English important for your future？)

4. 英語を聞くとき、英語のイントネーションに注意を向けますか。（以下に ｜ を入れてください）
   (When you listen to English, do you pay attention to intonation?)

   考えたこともない 少しは注意する とても注意して聞く
   |----------|----------|

5. 英語を聞くとき、英語の弱く話される箇所に注意を向けますか。（以下に ｜ を入れてください）
   (When you listen to English, do you pay attention to unstressed words?)

   考えたこともない 少しは注意する とても注意して聞く
   |----------|----------|

Appendix B

Summary of questions in the post-questionnaire

Q1: Did you enjoy the following activity?
1-1: dictation
1-2: group work
1-3: shadowing <DgS[a] group only>
Q2: Was the following activity easy for you?
2-1: dictation
2-2: shadowing <DgS group only>
Q3: Do you think the following activity was useful?
3-1: dictation for listening
3-2: dictation for speaking
3-3: shadowing for listening <DgS group only>
3-4: shadowing for speaking <DgS group only>
Q4: Would you recommend the following activity to English learners like you?
4-1-Dg: dictation and group work <Dg[b] group only>
4-1: dictation <DgS group only>
4-2: group work <DgS group only>
4-3: shadowing <DgS group only>

Note. [a]DgS and [b]Dg are the experimental groups: the assigned activities were dictation, group work, and shadowing for the DgS group, and dictation and group work for the Dg group.

Appendix C

Dictation Exercise #1 Class ( ) Group ( )

4. Excuse ( ), ( ) ( ) have ( ) shirt ( ) blue?
(A) Let ( ) check ( ) stock.
(B) ( ) shirt ( ) ( ) really well.
(C) ( ) love blue ( ).