The Effectiveness of Integrated Teaching Methods in English as a Foreign Language Classrooms

Mamiko ORII-AKITA[1]

Abstract

Reflecting that language skills function in tandem, *integrated* or *multi-skill instruction* has become a trend in language teaching. However, many studies evaluating this approach only report the results of student evaluations. Even empirical studies often employ problematic methods, and integrated approaches still require investigating in controlled classroom studies. This study, among intermediate English communication learners at a Japanese university, investigated the effectiveness of integrating listening comprehension, speaking, and pronunciation training. More specifically, three types of integrated instruction were carried out to examine which combination is the most effective way to improve learners' listening comprehension. The top-down group received explicit metacognitive strategy training, and the bottom-up group received *implicit* pronunciation instruction through dictation activity (with special focus on prosody). The interactive group received *explicit* pronunciation instruction (bottom-up training focusing on prosody), and also engaged in shadowing training. This group also received top-down explicit metacognitive strategy training. Three instructional groups — a bottom-up group, a top-down group, and an interactive group — underwent 20 class sessions using the same materials. The interactive group outperformed the other groups on a delayed posttest, indicating the superiority of an interactive approach when carrying out integrated or multi-skilled instruction.

1. Background

1.1 Introduction

On the basis of the understanding that in meaningful communication, people employ language skills not in isolation but in tandem, *integrated* or *multi-skill instruction* has emerged as a current trend in language teaching (Hinkel, 2006, p. 113). It has also been argued that integration of speaking and listening training is beneficial for language learners (Lynch, 2009). However, many studies report only the results of student evaluations as an indication of the integrated method's effectiveness (e.g., Edwards [2006]; Fujioka [2003]). Even when empirical studies have been carried out, their methods have often been problematic. For example, few studies have collected control group data to demonstrate the effectiveness of integrated teaching methods as compared to other methods. In addition, few studies have administered a delayed

posttest to assess whether participants' performance was sustained after the experiment. Thus, integrated approaches still need to be investigated in carefully controlled classroom studies including (1) multiple experimental groups, (2) an objective test of learners' listening-comprehension improvement, and (3) a delayed posttest.

This study investigated the effectiveness of an approach integrating three aspects of language instruction: *listening comprehension instruction* (top-down metacognitive strategies training), *pronunciation instruction* (bottom-up prosody training), and *speaking instruction* (shadowing training or dialogue practice). Three types of integrated instruction were carried out to examine which combination is the most effective way to improve learners' listening comprehension. The top-down group (TG) received metacognitive strategy training, and the bottom-up group (BG) received *implicit* pronunciation instruction through dictation activity (with special focus on prosody). The interactive group (IG) received *explicit* pronunciation instruction (bottom-up training focusing on prosody) and engaged in shadowing training. This group also received top-down explicit metacognitive strategy training. The results should shed light on how best to conduct integrated language teaching in a foreign-language setting.

1.2 The Integrated Teaching Approach

In meaningful communication, people employ language skills not in isolation but in tandem. This has led to the emergence of integrated or multi-skill instruction in language teaching (Hinkel, 2006, p. 113). Integrated teaching models for communication cover various combinations of skills. For example, speaking might be taught alongside writing and vocabulary, or pronunciation might be taught using activities that also improve speaking skills. Pronunciation instruction can also be integrated into speaking and listening instruction (see Hinkel [2006] for further discussion).

Here, we will focus on studies that look at teaching listening and oral skills. According to Lynch (2009, p. 110), conversation usually involves both speaking and listening, and so the distance between instruction in these two skills should be reduced. Bahns (1995, p. 537) identifies the importance of preparing students for real-life social interaction, which requires the development of both listening and speaking skills. Shudo and Harada (2009) also emphasize the importance of listening comprehension training including interaction between students in class.

A number of studies have investigated the effectiveness of integrated teaching for listening and speaking. Fujishiro, Hiramatsu, and Miyaji (2007) developed web-based listening comprehension materials for Japanese junior high school English learners. Two experimental groups participated in 10 hours of experimental class sessions. A web-based-training group received online listening comprehension training *blended* with classroom speaking activities, while a non-web-based-training group underwent conventional classroom teaching incorporating listening comprehension exercises and speaking activities. An in-house listening test showed that both groups improved their comprehension abilities equally.

Sasaki, Fujita, and Tanaka (1999), in a small-scale study, found that integrated listening training, writing training and speaking training effectively improved advanced adult learners' listening comprehension. The Program of Integrated Listening Training (PILT), which combines top-down (summary writing) and bottom-up (translation activity, spot dictation, group discussion) approaches, was used for training over three months with six advanced adult learners, who were given three different mock listening tests (TOEFL,

TOEIC, and grade 1 STEP tests) before and after the training. A significant difference between the pre- and post-tests was found for TOEIC and STEP, showing that PILT improved listening comprehension for these advanced learners. However, the study did not include a comparison group to confirm whether this method was superior to other integrated teaching.

Edwards (2006) and Fujioka (2003) advocate multi-skill instruction involving listening comprehension, speaking, and pronunciation instruction. Edwards (2006) discusses the important relationship of interdependence between listening and speaking in normal conversation, and argues that speaking practice and pronunciation exercises can be beneficial to learners, because if they can accurately produce target-language sounds they may be more likely to perceive those sounds when listening to other speakers. Fujioka (2003) argues that English classes should take an integrated approach allowing focus on individual skills as necessary, and that to improve their listening skills, students should receive conceptual instruction in the features of spoken English (e.g., stress patterns, contractions, liaisons) and engage in pronunciation practice. However, neither Edwards (2006) nor Fujioka (2003) provided qualitative data to support their views.

Ishikawa and Nishigaki (2005) are among the few who have provided quantitative data supporting integrated teaching of listening comprehension and pronunciation. In their study, 30 junior high school students studied authentic listening materials through computer-assisted language learning (CALL) and then received pronunciation and shadowing training. In-house listening tests were administered before and after the instruction, and showed that the participants significantly improved their listening ability. However, these authors did not use a comparison group to confirm the effectiveness of their method or administer a delayed posttest to see if the educational effect would be sustained over a longer period.

Thus, although several studies have argued that the integration of speaking and listening training is beneficial for language learners, many have done so on the basis of student evaluations alone, and even when empirical studies have been carried out, their methods of implementation have often been problematic: few have used control groups to confirm the effectiveness of the recommended teaching method compared to other methods, and few have administered a delayed posttest to assess whether changes in participants' performance were sustained. Thus, integrated approaches, which teach students how to regulate multiple skills, still need to be investigated in carefully controlled classroom studies that include (1) multiple experimental groups, (2) an objective test to measure learners' listening-comprehension improvement, and (3) a delayed posttest.

1.3 Three Listening Strategies

Previous research into how language users derive meaning from the sounds they hear has proposed three types of processing models: *bottom-up*, *top-down*, and *interactive*.

Early research proposed a bottom-up model in which listeners build understanding starting with the smallest salient units of the acoustic message, phonemes, which are combined into words, phrases, clauses, and sentences (Carrel, 1988; Flowerdew & Miller, 2005; Rost, 1990). This was thus a traditional model understanding communication as simply a means of transmitting information.

Others argue that effective listeners use top-down processing. In this model, listening is seen as an active process in which the listener must seek out necessary information, rather than passively analyzing speech as a series of sounds (Flowerdew and Miller, 2005, p. 26). Also referred to as the process-based

approach, this model presupposes the use of previous background knowledge to predict content (Ross, 1975; Rubin, 1994). Previous studies have shown that effective listeners use top-down processing more than bottom-up processing (Cross, 2009, 2011; Vandergrift, 1997, Vandergrift & Tafaghodtari, 2010).

Recent studies have suggested that listening comprehension comes from an interactive interplay of information at different levels, requiring both bottom-up and top-down processing (O'Malley & Chamot, 1990). Several researchers have argued that effective listeners most often utilize the interactive approach (Buck, 2001; Cleary, Holden, & Cooney, 2007; Field, 2008; O'Malley, Chamot, & Kupper, 1989; Vandergrift & Goh, 2012).

Vandergrift (2007) asserts that teaching second language (L2) listening comprehension should include both top-down and bottom-up training, and that the best methods of doing so should be investigated in carefully controlled classroom studies.

One of the few studies of this kind was conducted by Goh (2000), who recommended a combination of instruction in segmentation (a bottom-up process) and strategic training (a top-down process) to develop learners' listening comprehension. She identified real-time listening difficulties faced by a group of 40 Chinese English learners based on their introspective reports (diaries, retrospective verbalizations, etc.), and found that around half of these problems were perceptual processing problems arising from ineffective attention allocation and word-recognition failure. On this basis, she (p. 71) suggests a "direct strategy" in which teachers "help learners improve their listening comprehension directly by providing them with practice in perception of selected sounds, content words, [and] pronunciation of new words and intonation features, such as prominence and tones." However, she emphasized that perception practice alone is insufficient and that training to adopt an appropriate listening strategy is also needed.

In section 1.4, it will be shown that the interactive approach resembles natural human language processing and therefore yields better results than other approaches. Then, in section 2, the results of an experimental study evaluating interactive teaching will be reported. Specifically, the study compares the effectiveness of an approach combining listening comprehension instruction (top-down metacognitive strategy training), pronunciation instruction (bottom-up prosody training), and speaking instruction (shadowing training) with that of either bottom-up or top-down instruction alone.

1.4 Interactive Teaching and Human Cognitive Processing

This section will discuss the effectiveness of interactive teaching from the viewpoint of human cognitive processing. Few studies have discussed interactive teaching in the context of language acquisition. However, it is important to do so, as interactive teaching resembles natural human language processing and therefore yields better results than other teaching methods.

Moore (2012, p. 416) notes that successful auditory perception integrates bottom-up, auditory *sensory information* with top-down, multimodal *cognitive information*. Moore, Ferguson, Edmondson-Hones, Ratib, and Riley (2010) conducted a large-scale study of auditory perception in 6 to 11-year-old schoolchildren in the United Kingdom, finding that children with auditory processing disorder have significantly impaired top-down function.

Studies on working memory in autism have important implications for our understanding of cognitive processing for listening comprehension, in that they show that human data processing requires bidirectional

cooperative circular flow between top-down and bottom-up processing. According to Koshino (2008), people with autism generally rely on low-level cognitive processes. This makes them good at visuospatial processing, a largely low-level process, but not at high-level, linguistically rooted cognitive tasks (Frith, 1989). On this basis, Frith and colleagues proposed the "Weak Central Coherence Theory," suggesting that people with autism have difficulty making connections between low and high-level information (Frith 1989; Hill & Frith 2003).

In a brain-imaging study, Koshino, Carpenter, Minshew, Cherkassky, Keller, and Just (2005), pointed out that people with autism are likely to have weak functional connectivity between the frontal and occipital regions, and suggested that they lacked cooperative circular flow. On this basis, Koshino et al. argued that cooperative circular flow is essential for smooth processing of verbal information. In other words, it is necessary to process data bidirectionally instead of relying on one processing method only.

But how does bidirectional processing take place? The most widely held theory on the extraction of accumulated information is that the brain, like a computer, processes information relating to a single process and goal in parallel. That is, information search is not sequential or gradual but almost concurrent, with processes in mutual interaction. This led Marslen-Wilson (1987) to propose the "Cohort model" of speech/language processing, in which a few word candidates are selected based on the initial two or three phonemes of an input word, while concurrently, top-down processing using contextual information furthers understanding of the utterance.

Next, we will consider whence humans extract knowledge to process language in first language (L1) and second language (L2) listening. Long-term memory is divided into two broad classes according to the nature of the information: *declarative memory* and *procedural memory* (Squire, 1982, 1987). Declarative memory occurs with conscious awareness and can be consciously recalled. It is often further divided into two categories: *episodic memory* and *semantic memory*. Semantic memory stores general knowledge including facts, ideas, and linguistic knowledge, while episodic memory stores personal experiences tied to particular times and places. Procedural memory refers to implicit memories retained without conscious awareness, including the memory of routinized procedures and skills learned by using one's body.

In terms of listening processes, one difference between L1 and L2 listening is that in the latter, speech recognition and discrimination are conscious and slow. This difference stems from the different kinds of knowledge listeners turn to when processing L1 and L2 speech. In L1 listening, speech discrimination is instant and automatic because the language is the listener's native tongue; semantic memories are usually converted to procedural memories during acquisition, to be retrieved unconsciously whenever needed (Levelt, 1993).

By contrast, L2 listening is a conscious activity; one attempts to capture sound input while matching it with knowledge stored in semantic memory. Thus, processing takes more time than in L1 listening. It is therefore apparent that effective L2 listening requires (1) instant, automatic bottom-up processing for fast sound discrimination and (2) efficient top-down processing of contextual information.

Before discussing how to make L2 listening more like L1 listening, let us briefly review how humans process information.

According to Baddeley's working memory model (2000, 2002, 2007), information processing has a central executive that controls three subordinate systems: the *visuospatial sketchpad*, the *episodic buffer*,

and the *phonological loop*. The phonological loop is responsible for speech recognition; it temporarily stores speech input while the listener searches for necessary information in long-term memory. The phonological loop can be further divided into two subsystems: the *phonological short-term store* and *subvocal rehearsal*. Subvocal rehearsal is the conscious, active repetition of information. Baddeley further argues that subvocal rehearsal allows verbal information, which usually disappears from memory in approximately two seconds, to be retained longer in the phonological loop and that this function can elicit deeper levels of information processing, faster.

Next, let us discuss how to make L2 speech recognition faster and more automatic, that is, more like L1 listening. This entails improving the efficiency of bottom-up processing.

In this study, two methods of doing so were attempted. First, *prosody instruction* was used to enhance explicit phonological knowledge of the L2, inspired by the finding that EFL learners often lack sufficient L2 prosody knowledge to process English speech adequately (Akita, 2001; Orii-Akita[a], 2014; Kadota, 2012; Pennington, 1989, 1998;). Prosody instruction focuses explicitly on rhythm, stress, and sound change. Gilbert (1987) argues that prosody instruction is more effective than segment instruction (see also Akita, 2005, 2006; Akita-Orii, 2008; Pennington and Ellis, 2000; Ueno, 1998). Therefore, the author decided to give her interactive group prosody instruction rather than segment training — as their bottom-up training.

Following this, *shadowing* training was implemented. Shadowing is a technique often used for training interpreters: students immediately repeat the aural input as they listen to the CD. In this activity, subvocal rehearsal is used to retain the speech sounds in a phonological loop in working memory (Kadota, 2006, p. 14). The aim of this training was to gradually reduce the cognitive load required for subvocal rehearsal by enabling learners to process more information, (potentially) facilitating the automation of speech perception. According to Kadota (2012), in this situation, linguistic knowledge is likely to be converted from explicit declarative memory to implicit procedural memory. For more discussion on the effectiveness of shadowing training and its relationship to human cognitive processing, see Orii-Akita (2014)[a].

2. The Study

2.1 Overview of Experiment Design

The data for this study were collected during experimental class sessions conducted in three intermediate English communication classes taught by the author in 2012. The students in the experimental classes were from various departments, including Japanese, Geography, Natural Science, and Math. Twenty 90-minute experimental classes were conducted between April and November.

For these classes, students were assigned to suitable levels (basic, intermediate, advanced) according to results from the Web-based Test for English Communication (WeTEC), an evaluation tool used at the author's university and taken by over 10,000 people each year (including those who take it multiple times), based on CASEC (Computerized Assessment System for English Communication), a test provided by the Society for Testing English Proficiency (STEP), Tokyo. Students registered for the classes they were interested in according to their assigned proficiency level. None of them had lived or studied abroad longer than a month. There were 30 students in each class, but students meeting one or more of the following conditions were excluded from analysis: students who were absent on seven or more occasions (missing more

than 30% of classes, and thus failing to qualify to earn class credit); students who didn't complete all three WeTEC assessments; and students majoring in English language and literature (because they take English phonetics as a first-year requirement and thus already have some knowledge of English phonetics and phonology). Informed consent was obtained at the first session. The participants were informed that personally identifiable information would be removed from the dataset and that the data would be used only for this analysis and to provide their course grades.

Participants' listening abilities were measured three times using the listening sections of WeTEC. This tool uses an Item Response Theory (IRT)-based computerized adaptive testing (CAT) system and adapts questions to the student's individual level. The whole test last approximately one hour, with 45 listening questions and 50 question with reading and vocabulary section. It is considered a reliable assessment, with a high correlation (r=0.858) with students' TOEIC (Test of English for International Communication) scores (Nakano, Owada, Ueda, Oya, Tsutsui, Kondo, & Yoshida, 2012, p. 54).

The entire WeTEC test (four sections) was administered; however, since this study aims specifically to investigate the relationship between teaching method and listening ability, the author analyzed only the listening sections (Sections 3 and 4). The test provides scores listening for gist (Section 3) and listening for details (Section 4). This approach is in line with the top-down teaching model, which is one reason WeTEC was adopted over other online tests.

A pretest was conducted before the first experimental session and a posttest administered immediately after the final session, in November 2012. In addition, a delayed posttest was implemented two months after the end of the experimental classes, in February 2013, to discover whether the educational effect was sustained.

2.2 Experimental Classes and Lesson Format

Overview. Each group underwent group-specific listening training using the same textbook, developed by the author. The bottom-up group (n = 17; henceforth, BG) completed the dictation activity, while the top-down (n = 20; henceforth, TG) and interactive (n = 18; henceforth, IG) groups completed listening strategy exercises. Listening comprehension activities took approximately 60 of the 90 minutes per class. The participants also engaged in approximately 30 minutes of speaking activities; in these, BG and TG produced and presented dialogues in pairs, while IG engaged in basic prosody-focused pronunciation instruction and practice, followed by a shadowing activity.

Listening training lesson format. A listening textbook and CD (on travel) developed by the author were used in all three groups. Dialogues from the CD were played five times for each group. Most previous studies have used different teaching materials for each experimental group; few studies have provided uniform input across groups because of the practical difficulties. In contrast, in this study the author used the same listening materials for each group, aiming thereby to measure the educational effect of each teaching method with greater accuracy. This methodology has not yet been implemented in long-term studies.

BG engaged in dictation activities designed using the bottom-up model. These activities *implicitly* directed students' attention to sound changes. The blanks in the dictation activity represented clear examples of the prosodic point being covered, as well as key expressions. Afterward, the instructor distributed the full script and explained the vocabulary and grammar.

TG received *explicit* instruction in listening strategies designed using the top-down model. Following Vandergrift (1997), eleven strategies were chosen for the participants to receive instruction in:

(1) *directed attention*;
(2) *selective attention*;
(3) *verifying predictions*;
(4) *linguistic inferencing*;
(5) *voice inferencing*;
(6) *inferencing between parts*;
(7) *personal elaboration*;
(8) *world elaboration*;
(9) *questioning elaboration*;
(10) *note-taking*; and
(11) *cooperation*.

Listening strategy instruction included the following three activities: (1) *pre-listening* to activate learners' schemata on the topic; (2) *first listening*, with exercises, for overall understanding; and (3) *second listening*, with exercises, for specific information (as recommended in Ellsworth, [2006] and Wilson [2008]). At times, a listening strategy handout (distributed at the start of the course) was used to explicitly teach students how to scan for essential information, predict the content and intended meaning of the story, and utilize logical and visual information.

IG received instruction on the same listening strategies and completed the same listening exercises as TG.

Speaking activity lesson format. BG and TG were tasked with creating dialogues and sentences about the unit topic, either in pairs or individually, and presenting them in small groups or to the class as a whole. For example, unit 5 covered "Ordering at a restaurant", and students thus practiced ordering food. The production activity lasted about 30 minutes.

As for IG, participants were given prosody instruction and engaged in pronunciation practice for approximately 30 minutes. The first 10 minutes were devoted to outlining the main features of the new pronunciation point in simple, jargon-free terms. For example, units 1 and 2 focused on the rhythm of English and the use of weak forms. Subsequent units covered thought groups, assimilation, linking, and elision, with a new pronunciation point introduced roughly every other unit.

The last 20 minutes were spent on practical pronunciation exercises. First, the class practiced pronunciation using very simple sentences (e.g., the phrase "Would you hurry up?" was used to practice assimilation). Then, they did pronunciation exercises using the listening script, in pairs. The relevant prosodic points were identified either by the students or by the instructor, who demonstrated them and asked the students to repeat them individually. Finally, the students worked on pronunciation exercises drawn from the spoken material. The dialogues were designed to reinforce the prosodic points covered in the unit — at least 10, but if possible up to 20, instances in each (each dialogue lasted 30–45 seconds). From units 14 to 20, instead of pronunciation training through oral reading, participants did a shadowing activity.

2.3 Results

Overview of data analysis. We conducted a two-way repeated-measures ANOVA of students' scores on Section 3 (listening for overall understanding), Section 4 (listening for specific information), and a combined total (Section 3 + Section 4). For each set of scores, the within-subjects factor of *test* (3 levels: pretest, posttest, and delayed) and the between-subjects factor of *group* (3 levels: BG, TG, and IG) were tested at the significance level of 5%. Mauchly's sphericity test for homogeneity of variance, a prerequisite for repeated-measures ANOVA, was done for each section, and indicated that the assumption of sphericity had not been violated for Section 3 ($p = 0.469$ [< 0.05; n.s.]), Section 4 ($p = 0.389$ [n.s.]) or total score ($p = 0.992$ [n.s.]). The general linear model (GLM) in SPSS was carried out to calculate the repeated-measures ANOVA.

Table 1 *Summary of ANOVA results*

Section	Test	Main effect				Interaction (group x test)	
		Group		Test			
		F	*p*	*F*	*p*	*F*	*p*
Section 3	Pretest	2.088	0.134	13.477	0.000***	5.830	0.000***
	Posttest						
	Delayed						
Section 4	Pretest	0.702	0.500	9.065	0.000***	3.183	0.016*
	Posttest						
	Delayed						
Total score	Pretest	0.622	0.541	28.308	0.000***	10.288	0.000***
	Posttest						
	Delayed						

Note. *p $<$ 0.05, *p $<$ 0.01, ***p $<$ 0.001

Table 2 *Descriptive statistics*

Section	Test	Group					
		BG ($n = 17$)		TG ($n = 20$)		IG ($n = 18$)	
		Mean	SD	Mean	SD	Mean	SD
Section 3	Pretest	165.65	22.366	159.40	24.659	163.22	12.959
	Posttest	171.71	14.225	166.50	21.309	173.17	16.078
	Delayed	166.29	21.502	169.15	23.703	190.72	13.123
Section 4	Pretest	148.00	17.346	144.80	19.606	134.33	14.000
	Posttest	153.94	19.986	146.60	15.823	148.61	19.135
	Delayed	151.47	20.100	147.90	18.046	153.17	16.104
Total score	Pretest	313.65	33.386	304.20	38.164	297.56	20.171
	Posttest	325.65	26.907	313.10	31.616	321.78	29.883
	Delayed	317.76	32.128	317.05	35.685	343.89	21.884

Tab1e 3 *Summary of post-hoc ana1ysis (mu1tip1e comparison)*

Section	Post-hoc analysis (comparison of groups by test): Multiple comparison (Tukey's HSD test)					Post-hoc analysis (comparison of tests by group): Bonferroni-adjusted paired t-test				
	Test	Group (l)	Group (J)	Difference in means (l-J)	p value	Group	Test (I)	Test (J)	Difference in means (l-J)	p value
Sections 3	Pretest	BG	TG	6.25	0.635	BG	Pretest	Posttest	−6.06	0.578
		BG	IG	2.42	0.936		Pretest	Delayed	−0.65	1.000
		TG	IG	−3.82	0.838		Posttest	Delayed	5.41	0.436
	Posttest	BG	TG	5.21	0.647	TG	Pretest	Posttest	−7.10	0.447
		BG	IG	−1.46	0.968		Pretest	Delayed	−9.75	0.139
		TG	IG	−6.67	0.482		Posttest	Delayed	−2.65	1.000
	Delayed	BG	TG	−2.86	0.903	IG	Pretest	Posttest	−9.94	0.123
		BG	IG	−24.43	0.002*		Pretest	Delayed	−27.50	0.000***
		TG	IG	−21.57	0.005*		Posttest	Delayed	−17.56	0.000***
Section 4	Pretest	BG	TG	3.20	0.840	BG	Pretest	Posttest	5.94	0.307
		BG	IG	13.67	0.058		Pretest	Delayed	−3.47	0.820
		TG	IG	10.47	0.158		Posttest	Delayed	2.47	1.000
	Posttest	BG	TG	7.34	0.448	TG	Pretest	Posttest	1.80	1.000
		BG	IG	5.33	0.666		Pretest	Delayed	−3.10	1.000
		TG	IG	−2.01	0.939		Posttest	Delayed	−1.30	1.000
	Delayed	BG	TG	3.57	0.822	IG	Pretest	Posttest	−14.28	0.010*
		BG	IG	1.70	0.959		Pretest	Delayed	18.83	0.000***
		TG	IG	−5.27	0.646		Posttest	Delayed	−4.56	0.787
Total score	Pretest	BG	TG	9.45	0.642	BG	Pretest	Posttest	−12.00	0.004**
		BG	IG	16.09	0.300		Pretest	Delayed	−4.12	1.000
		TG	IG	6.64	0.797		Posttest	Delayed	7.88	0.2 50
	Posttest	BG	TG	12.55	0.411	TG	Pretest	Posttest	−8.90	0.4 56
		BG	IG	3.S7	0.92 1		Pretest	Delayed	−12.85	0.051
		TG	IG	−S.68	0.642		Posttest	Delayed	−3.95	1.000
	Delayed	BG	TG	0.71	0.997	IG	Pretest	Posttest	−24.22	0.000***
		BG	IG	−26.12	0.039*		Pretest	Delayed	−46.33	0.000***
		TG	IG	−26.84	0.025*		Posttest	Delayed	−22.11	0.000***

Note. *p < 0.05, *p < 0.01, ***p < 0.001

In cases of significant interaction, a multiple comparison of groups by test (using Tukey's HSD test) and of tests by group (using Bonferroni-adjusted paired t-tests) was done. Tables 1–3 show the results.

Results for Section 3. The main finding of the within-subjects test was a significant difference in mean values between the three tests (F[2,104] = 13.477, p = 0.000, partial η^2 = 0.206). The significant interaction between group and test (F[4,104] = 5.830, p = 0.000, partial η^2 = 0.183) indicates an interaction, and thus that the changes in test scores between levels differ between groups.

Next, a between-subjects analysis (paired samples) was carried out. The main finding was that there

was no significant difference in mean score between groups (F[2,52] = 2.088, p = 0.134 [n.s.], partial η^2 = 0.074). However, as stated above, there was a significant interaction between group and test, showing that the changes in mean test score between levels differed between groups.

Further, also as mentioned, there was a significant interaction between group and test. This shows that the changes in mean test score by level differ by group. The results here show no significant difference between groups; however, since we knew that an interaction was present, multiple comparison of groups by test was done.

Tukey's HSD test results (multiple comparison of groups by test) show a difference between the means of BG and IG (p = 0.002) and of TG and IG (p = 0.005) on the delayed test, with IG significantly higher on Section 3 than BG or TG.

Next, a multiple comparison (Bonferroni-adjusted paired t-test) of tests by group was done. Differences were found between the pretest and delayed (p = 0.000) and posttest and delayed (p = 0.000) means for IG, with Section 3 scores significantly higher on the delayed test than on the others.

Here follows a summary of the results for Section 3 (see also Figure 1). The main results showed that test (pretest, posttest, delayed) was a significant factor affecting score changes in Section 3, while group (BG, TG, IG) was not. However, a significant interaction was identified between group and test, showing that score changes differed depending on the combination of levels for each factor. Therefore, as a post-hoc analysis, we carried out a multiple comparison for each level. It showed a significant difference in scores on the delayed test between BG and IG and between TG and IG (with IG scoring significantly higher in both cases). This shows the effectiveness of the interactive training received by IG. Further, the results of the multiple comparison tests by group were noteworthy: although IG showed no significant improvement from pretest to posttest, they did show a significant improvement in the delayed posttest, showing that interactive training continues to have an effect on students' scores even after completion of training (i.e.,

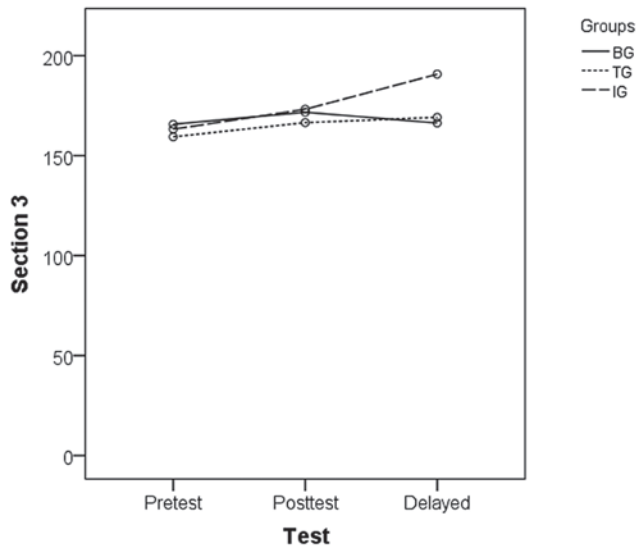

Figure 1. Mean values for Section 3.

after the end of the experimental study).

The scores for BG and TG, however, were somewhat stagnant. We provided listening strategy training to TG as a way of teaching listening for overall understanding, the skill measured in Section 3. It is unfortunate from an educational standpoint that TG showed no improvement in this area.

Results for Section 4. The main findings from the ANOVA for the within-subjects factor showed a significant difference in mean values between the three tests (F[2,104] = 9.065, p = 0.000, partial η^2 = 0.148). The interaction between group and test was also significant (F[4,104] = 3.183, p = 0.016, partial η^2 = 0.109), showing that changes in test scores by level were different by group.

Next, an ANOVA for the between-subjects factor (groups) was conducted. No significant differences were observed between groups (F[2, 52] = 0.702, p = 0.500 [n.s.], partial η^2 = 0.026). However, a significant interaction was observed between group and test, suggesting that changes in mean test scores by level were different by group.

A post-hoc analysis was done to assess this interaction. First, multiple comparison results for the groups showed that for any test, there were no differences in mean values between any two groups.

The multiple comparison results show a significant difference between IG's pretest and posttest scores (p = 0.010) and pretest and delayed scores (p = 0.000). IG improved their mean score on the posttest, and the improvement remained on the posttest.

Regarding the factors explaining these score changes on Section 4, the main finding was a significant difference between the three tests (pretest, posttest, and delayed) but not between the groups (BG, TG, and IG). In addition, the interaction (group × test) was significant, showing that the different combinations of scores for each factor resulted in different patterns of change. Thus, multiple comparison results showed no significant difference between group scores, no matter the test. Conversely, when comparing tests across groups, a significant difference was observed for IG (only): posttest and delayed scores were higher than pretest scores.

As in Section 3 (listening for overall understanding), then, IG scores improved in Section 4 (listening for specific information; Figure 2). In Section 3, IG showed a significant improvement on the delayed posttest, while in Section 4, the improvement came on the (immediate) posttest and was maintained on the delayed posttest. Conversely, even though TG was provided with listening strategies and BG participated in dictation activities, neither group recorded improved scores on the listening test, and so it appears that these teaching approaches had no lasting effect.

Results for total score. We first tested within-subjects effects. The main effect of test was significant (F[2,104] = 28.308, p = 0.000, partial η^2 = 0.352), indicating mean score differences among the three tests. In addition, test scores trended differently by group (BG, TG, and IG) due to the significant interaction (test × group) observed (F[4,104] = 10.288, p = 0.000, partial η^2 = 0.284).

Next, an ANOVA was performed on the between-subjects factor (group). The main effect of group was not significant (F[2,52] = 0.622, p = 0.541 (n.s.), partial η^2 = 0.023), indicating that there were no mean differences among the three groups.

Given that the interaction was significant, post-hoc analyses were conducted. Multiple comparison results by test showed that the mean differences on the delayed test between BG and IG and between TG and IG were significant (p = 0.039 and p = 0.025, respectively). Scores were significantly higher for IG.

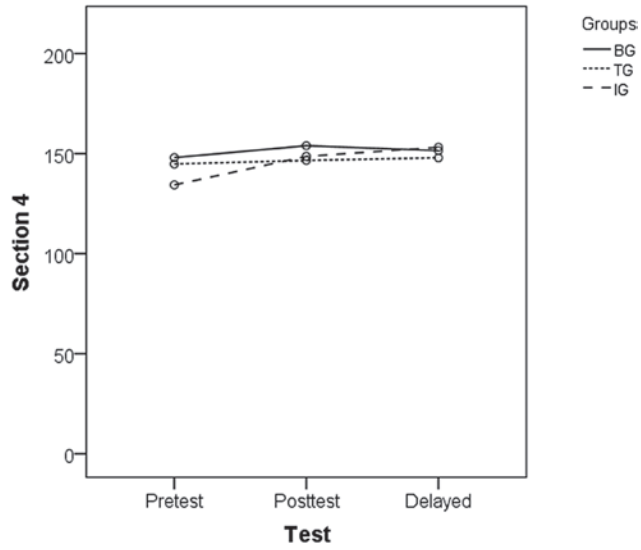

Figure 2. Mean values for Section 4.

Next, multiple comparison for each test by group indicated a significant difference between pretest and posttest within BG (p = 0.004): the posttest average was significantly higher. Meanwhile, mean differences between the following pairs of tests were observed within IG: pretest vs. posttest (p = 0.000), pretest vs. delayed (p = 0.000), and posttest vs. delayed (p = 0.000). For this group, scores improved over time, from pretest to posttest and then from posttest to delayed test.

Thus, although test (pretest, posttest, and delayed) turned out to be a significant factor for changes in total scores, group (BG, TG, and IG) was not a significant factor. Furthermore, the interaction (group × test) was significant, indicating that score changes differed according to the different combinations of levels of each factor. Post-hoc analyses for each level showed that only the delayed test was significant between groups by test, specifically for BG vs. IG and for TG vs. IG. IG scores were significantly higher in both cases.

Meanwhile, as can be seen in Figure 3, the comparison of tests by group indicated a significant difference between pretest and posttest in BG (the posttest score was significantly higher), while scores in the IG group significantly improved from pretest to posttest and from posttest to delayed test. The fact that the average score increased significantly over the course of the three tests indicates that IG not only remained effective after two months, but also spurred further improvement in listening skills. As for BG, we can see that while the mean score improved from pretest to posttest, the mean on the delayed test returned to the same level as the pretest, indicating that the effect of the experimental class was only temporary for this group. Unfortunately, TG was slow to improve, showing no change in mean score.

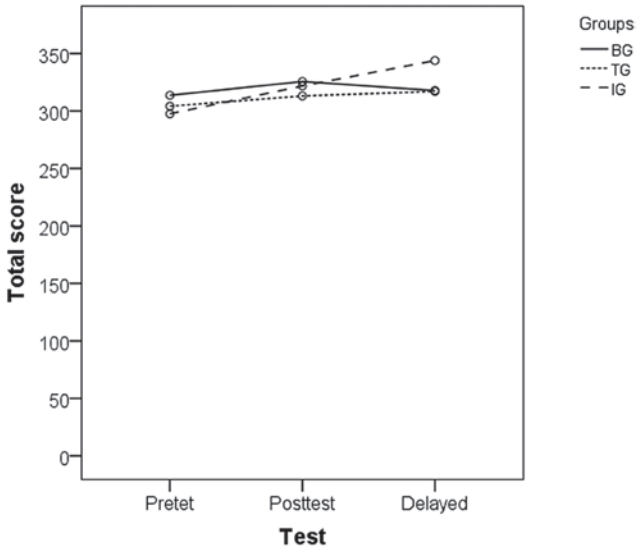

Figure 3. Mean values for total score.

3. Discussion

Summary and examination of results. This study investigated the effectiveness of integrating listening comprehension, speaking, and pronunciation training. More specifically, three types of integrated instruction were carried out to examine which combination is the most effective way to improve learners' listening comprehension.

For BG, no improvement in mean score was observed on either Section 3 or Section 4. However, their posttest mean for total score was higher than their pretest mean. That is, even though neither BG's score for Section 3 nor that for Section 4 increased significantly, when the scores for both sections were added together, the increase did become significant. However, despite this improvement in total score on the posttest, BG's delayed posttest score returned to the level of the pretest. In other words, the effect of the experimental instruction was only temporary in this group.

Unfortunately, scores for TG did not improve on Section 3, Section 4, or for total score, showing that the teaching of listening strategies, implemented as a single method, did not contribute to any improvement in listening abilities for TG. In particular, the goal of training the students in listening for overall understanding by teaching them about various listening strategies seems not to have been achieved, as TG's score for listening for overall understanding (Section 3) showed no improvement.

However, this conclusion still seems premature given that in 2011, the author (with a colleague) collected listening data (published as Akita-Orii & Oga [2012]) with a similar methodology to the current paper. The major difference is that those data were collected on the basis of 14 class sessions (conducted over the period of 3 months) instead of 20. In the earlier study, total posttest scores for both IG and TG improved, indicating that the teaching of listening strategies was effective in TG as well as IG. Nevertheless, the data collected for the current study do appear to show that instruction in top-down listening alone had no effect, even with the larger number of classes.

To try to determine why no effect was observed in TG in the current study, the author referred, in addition to the WeTEC results, to those of the regular end-of-term examinations conducted in each class. These examinations were administered by the author on two occasions, at the end of terms 1 (August 2012) and 2 (February 2013). They included a listening test on the material covered as well as written grammar and vocabulary tests, along with questions requiring the students to translate Japanese text into English. On these exams, results for TG were lower than for BG or IG, seemingly meaning that for whatever reason, motivation was low in this group. However, the author did not examine the psychological state of the participants and thus cannot say for certain that low motivation caused TG's low performance. Future research should consider the psychological state of study participants, for example by surveying them to ascertain their motivation. Whatever the reason for these poor results in TG, the disappointing but logical conclusion is that the experimental instruction implemented for this study likely had no permanent or temporary educational effect on this group.

In contrast, significant improvement was observed on Sections 3 and 4 and for total score in IG, which received both top-down (listening strategies) and bottom-up (prosody) teaching. In Section 3, which measured listening for overall understanding, no significant improvement was observed on the immediate posttest, but significant improvement was observed on the delayed posttest. This shows that interactive teaching improved students' scores even after it had concluded (i.e., after the listening and speaking training was finished). A comparison between groups for each test showed significant differences in delayed posttest scores between BG and IG and between TG and IG.

In Section 4, which measured listening for specific information, improved scores were observed in the immediate posttest and maintained in the delayed posttest. However, the score of IG on Section 4 was not significantly higher than those of TG and BG.

With regard to total scores, however, IG improved significantly in the posttest and again in the delayed posttest, showing that the effect not only persisted but also grew over time. Therefore, the data show that interactive teaching was an effective teaching method in conjunction with the classes used in this experiment.

The data collected in 2011 (Akita-Orii & Oga, 2012) will now be compared with those presented in this paper. In the 2011 data, TG and IG both showed identical overall improvement, meaning that IG's score was unexceptional. There are (at least) three plausible reasons why interactive teaching did not produce significant effects in the 2011 data. (1) The prosody teaching approach may have relied excessively on oral explanation, which the students may not have retained well in mind. (2) In the prosody teaching, a recorded native sound source was not used. Instead, short sentences were read by the author, and thus students may not have had enough chance to get used to the prosodic structure of English utterances, both because the sentences may have been too short and simple to provide quantitatively adequate data or a representative range of prosodic phenomena, and also perhaps because of issues regarding the author's non-native pronunciation. (3) As Akita (2001) has noted, it takes learners a long time (roughly at least half a year) for the learners to be able to start to perceive and produce correct prosodic structure of a target language, even in the environment that the target language is spoken on the daily basis. This may also apply to listening-comprehension abilities, and thus the number of listening-teaching classes administered in the 2011 studies may have been insufficient.

These issues were addressed by the following methodological improvements in the present study: (1) prosody instruction incorporated note-taking, (2) pronunciation practice used a native sound source and shadowing activities to help students become used to English prosodic structures, and (3) the number of classes was increased to 20. These changes are likely why listening ability improved more in IG than in TG or BG. The possibility does still need to be considered that IG outperformed TG only because the score obtained by TG was stigmatized. However, as a significant improvement was observed for IG on both Sections 3 and 4 as well as for total score, it seems instead more likely that greater educational effects were achieved through interactive teaching in the present study than previously.

Implications and limitations of the research. As discussed in Hinkel (2006), the multi-skill or integrated approach to instruction has been a notable trend in language teaching for some time. However, its effectiveness has not been examined sufficiently. Although it has been used in many studies assessing student learning on the basis of student evaluations, this alone does not offer satisfactory evidence of the approach's effectiveness. In addition, it need to be investigated in carefully controlled classroom studies, namely a longitudinal qualitative study using objective tests rather than relying on student evaluations alone and also, as in this study, incorporating comparison groups and a delayed posttest, which will establish whether the effect is temporary or permanent and if the latter, whether it remains steady or increases further over time, and also cover the possibility that there is no immediate effect but an aftereffect at delayed posttest, as in this study.

Although the findings presented here suggest that integrated teaching has the potential to significantly impact L2 listening, further research is needed not only to confirm the findings but also to extend them further. To begin with, while objective tests like those implemented here are beneficial for measuring listening comprehension ability, they do not reveal whether the improvement is because the participants have learned to use the taught strategies in listening comprehension. This question needs to be examined in future studies, through questionnaires or the introspective method.

Furthermore, in addition to the three experimental groups implemented in this study, the use of another bottom-up group, receiving explicit prosody instruction, will make collected data more informative, because we will be able to examine whether the effect differs when bottom-up teaching is carried out implicitly (as in a dictation activity) or explicitly (as in prosody instruction for IG).

In addition, future studies should examine the participants' psychological state. Assessments (evaluations and questionnaires) should be implemented multiple times: before, during, immediately after, and a couple of months after the experimental classes. This will enable the researcher to observe changes in psychological state among participants and may help in interpreting the main results.

There is another reason to implement student assessments: there may be a set span within which students can retain a high level of motivation when engaging in similar listening tasks. Instructors and researchers tend to believe that more instruction (more classes) has a positive effect on learners. However, attending too many classes with similar activities may bore learners and their motivation level may decline, resulting in stagnant listening scores. In fact, this may explain the low score of TG in the present study to some extent, since the present participants attended more experimental classes than the 2011 participants. In comparison, IG, which underwent identical listening strategy training to TG, may have been helped to maintain a fresh attitude and strong motivation toward the material by the introduction of the shadowing

activity in the 14th session.

In short, future research employing improved methods of data collection will help researchers examine the effectiveness of an integrated approach to teaching English as a foreign language.

4. Conclusion

The objective of this research was to examine the effectiveness of an integrated teaching approach for the teaching of EFL at the university level. This study investigated the effectiveness of integrating listening comprehension, speaking, and pronunciation training. More specifically, three types of integrated instruction were carried out to examine which combination is the most effective way to improve learners' listening comprehension.

As compared to the single-method learners (i.e., TG and BG), the group undergoing interactive instruction learned to more smoothly employ parallel bottom-up and top-down processing, due to the tandem combination of practice in the former and instruction in listening strategies to encourage the latter. Thus, from the viewpoint of human cognitive processing, interactive teaching — concurrent provision of bottom-up and top-down practice — contributes to the improvement of listening skills more than either of these methods alone.

[Notes]

1 This work was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (C) [24520660]. Results of the study were presented at New Sounds 2013 (Concordia University), and the manuscript was later published in Concordia Working Papers in Applied Linguistics. This paper is an expansion and re-analysis of Orii-Akita (2014)[b].

[References]

Akita, M. (2001). *The Phonological Development of Adult Japanese Learners of English: a longitudinal study of perception and production.* Ph.D. thesis. University of Durham, U.K.

Akita, M. (2005). The effectiveness of a prosody-oriented approach in L2 perception and production. In A. Brugos, M. R. Clark-Cotton, & S. Ha. (Eds.) *Proceedings of the 29th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 24–36.

Akita, M. (2006). Global Foreign Accent and the effectiveness of the prosody-oriented approach. In T. Magnitskaia & C. Zaller. (Eds.) *Proceedings of the 30th Annual BUCLD.* Somerville, MA: Cascadilla Press, 1–14.

Akita-Orii, M. (2008). Accentedness judgment and L2 phonology in EFL classrooms. *Journal of English Phonetics* (11・12), 265–280.

Akita-Orii, M. & Oga, K. (2012). A Comparison of three different methods of teaching listening comprehension. *Journal of English Phonetics* (16), 147–158.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Science*, *4*(11), 417–423.

Baddeley, A. D. (2002). Is working memory still working? *European Psychologist*, *7*(2), 85–97.

Baddeley, A. D. (2007). *Working Memory, Thought, and Action.* Oxford: Oxford University Press.

Bahns, J. (1995). Retrospective review article: There's more to listening than meets the ear. *System, 23*(4), 531–537.

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

Carrel, P. (1988). Some causes of text-boundness and schema interference in ESL reading. In P. Carrel, J. Devine, & D. Eskey (Eds.), *Interactive Approaches to Second Language Reading* (pp. 101–113). New York: Cambridge University Press.

Cleary, C., Holden, B., & Cooney, T. (2007). *Top-up Listening.* Tokyo: Abax ELT Publishers.

Cross, J. (2009). Effects of listening strategy instruction on news videotext comprehension. *Language Teaching Research, 13*(2), 151–176.

Cross, J. (2011). Metacognitive instruction for helping less-skilled listeners. *ELT Journal, 65*(4), 408–416.

Edwards, N. (2006). Approaches to teaching the language skill of listening: Practical and theoretical considerations. *English and English-American Literature, 41*, 121–132.

Ellsworth, I. E. (2006). Developing an English listening course for low proficiency Japanese university students. *The Journal of Chiba University of Commerce, 43*(4), 77–91.

Field, J. (2008). *Listening in the Language Classroom*. Cambridge: Cambridge University Press.

Frith, U. (1989). A new look at language and communication in autism. *British Journal of Disorders of Communication*, 24, 123–150.

Fujioka, M. (2003). Language learning as an integrated process: Inter-relations among listening, speaking, reading, and writing. *Kinki University Department of Language Education Bulletin, 2*(2), 31–37.

Fujishiro, N., Hiramatsu, S., & Miyaji, I. (2007). Effects of communication practice utilizing the digital learning materials based on WBT for English listening in the blended learning approach. *Japan Society for Educational Technology, 31*, 85–88.

Flowerdew, J., & Miller, L. (2005). *Second Language Listening.* Cambridge: Cambridge University Press.

Gilbert, J. (1987). Pronunciation and listening comprehension. In J. Morley (Ed.), *Current Perspective on Pronunciation* (pp. 33–39). Washington, D. C.: Teachers of English to Speakers of Other Languages.

Goh, C. C. M. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, 28, 55–75.

Hill, E. L., & Frith, U. (2003). Understanding autism: Insights from mind and brain. *Philosophical Transactions of the Royal Society of London,* 358, 281–289.

Hinkel, E. (2006). Current perspectives on teaching the four skills. *TESOL Quarterly, 40*(1), 109–131.

Ishikawa, M., & Nishigaki, C. (2005). A study of instruction aimed at improving English speaking ability among Japanese junior high school students (II. humanities and social sciences). *Bulletin of the Faculty of Education, Chiba University, 53*, 285–291.

Kadota, S. (2006). *Reading and Phonological Processes in English as a Second Language.* Tokyo: Kuroshio Shuppan.

Kadota, S. (2012). *Shadowing, Reading Aloud, and Science of English Acquisition.* Tokyo: Cosmopier.

Koshino, H. (2008). Working memory in autism (Jiheisho no waakingu memorii). In N. Osaka (Ed.), *Neural correlates of working memory* (pp. 161–192). Kyoko: Kyoto University Press.

Koshino, H., Carpenter, P. A., Minshew, N. J., Cherkassky, V. L., Keller, T. A., & Just, M. A. (2005). Functional connectivity in an fMRI working memory task in high-functioning autism. *NeuroImage*, 24, 810–820.

Levelt, W. J. (1993). The architecture of normal spoken language use. In G. Blanken, J. Dittman, H. Grimm, J. C. Marshall, & C.-W. Wallesch (Eds.), *Linguistic Disorders and Pathologies: An International Handbook* (pp. 1–15). Berlin: Walter de Gruyter.

Lynch, T. (2009). *Teaching Second Language Listening*. Oxford: Oxford University Press.

Marslen-Wilson, W. (1987). Functional parallelism in spoken word recognition. *Cognition*, *25*, 71–102.

Moore, D. R, Ferguson, M. A., Edmondson-Hones, A. M., Ratib, S., & Riley, A. (2010). The nature of auditory processing disorder in children. *Pediatrics*, 126, 382–390.

Moore, D. R. (2012). Listening difficulties in children: Bottom-up and top-down contributions. *Journal of Communication Disorders,* 411–418.

Nakano, M., Owada, K. Ueda, T., Oya, M. Tsutsui, E., Kondo, Y., Yoshida, S. (2012). JACET-Kanto Journal 8, 48–59.

O'Malley, J. M., & Chamot, A. U. (1990). *Learning Strategies in Second Language Acquisition.* Cambridge: Cambridge University Press.

O'Malley, J. M., Chamot, A. U., & Kupper, L. (1989). Listening comprehension strategies in second language acquisition. *Applied Linguistics*, *10*(4), 418–37.

Orii-Akita, M. (2014)[a]. Human cognitive processing and the interactive teaching method in EFL listening comprehension. *Gakujutsu Kenkyu, English Language and Literature,* Waseda University, vol. 62, 1–16.

Orii-Akita,M. (2014)[b]. The effectiveness of interactive teaching methods in EFL classrooms: a comparison with bottom-up and top-down methods. *Concordia Working Papers in Applied Linguistics*, vol. 5, 463–477.

Pennington, M. C. (1989). Teaching pronunciation from the top down. *RELC Journal*, 20, 207–225.

Pennington, M. C. (1998). The teachability of phonology in adulthood: A re-examination. *International Review of Applied Linguistics* 36, 323–341.

Pennington, M. C., & Ellis, N. C. (2000). Cantonese speakers' memory for English sentences with prosodic cues. *The Modern Language Journal*, *84*(3), 372–389.

Ross, R. N. (1975). Ellipsis and the structure of expectation. *San Jose State Occasional Papers in Linguistics*, 1, 183–91.

Rost, M. (1990). *Listening in Language Learning.* Essex: Longman.

Rubin, J. (1994). A review of second language listening comprehension research. *The Modern Language Journal*, 78(2), 199–221.

Sasaki, M., Fujita, N., & Tanaka, M. (1999). Effects of PILT (Program of Integrated Listening Training) on listening skills of advanced adult learners. *Akita English Studies, 41*, 25–35.

Shudo, S., & Harada, Y. (2009). Designing a syllabus for integrated language activities. *Humanitas, 47*, 1–11.

Squire, L. R. (1987). *Memory and Brain.* New York: Oxford University Press.

Squire, L. R. (1982). The neuropsychology of human memory. *Annual Review of Neuroscience.* 5, 241–73.

Ueno, N. (1998). *Teaching English Pronunciation to Japanese English Majors: A comparison of a suprasegmental-oriented and a segmental-oriented teaching approach.* Tokyo: Liber Press.

Vandergrift, L. (1997). The comprehension strategies of second language (French) listeners: A descriptive study. *Foreign Language Annals*, *30*(3), 387–409.

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, 40, 191–210.

Vandergrift, L., & Tafaghodtari, M. H. (2010). Teaching L2 learners how to listen does make a difference: An empirical study. *Language Learning, 60*(2), 470–497.

Vandergrift, L., & Goh, C. (2012). *Teaching and Learning Second Language Listening: Metacognition in action*. New York: Routledge.

Wilson, J. J. (2008). *How to Teach Listening*. Harlow, UK: Pearson Longman.