The Mismatch between Educational Policy and Classroom Practice: EFL Teachers' Perspectives on washback in Japan

Rika Tsushima

Department of Integrated Studies in Education
McGill University

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of
Master of Arts

Abstract

Research has reported that the occurrence of washback effects—the impact of high-stakes exams on classroom practice and activity–is often partly attributed to teachers (Cheng, 1997, 2000; Turner, 2001, 2006, 2009). In Japanese secondary schools, due to the strong influence of university entrance exams nationally, it is often argued that in practice, speaking-focused courses entitled *Oral Communication* (OC) do not focus on interactive language learning activities but on grammar exercises to prepare students for the high-stakes exams (Kikuchi & Browne, 2009). Using mixed methods, this thesis examines the current status of OC courses in relation to the national educational policy from the perspective of Japanese teachers of English as a foreign language (EFL). Quantitative data from a teacher survey ($N = 87$), qualitative data from classroom observations and a term exam analysis revealed that washback effects were more evident in the assessment practices of the courses than in teaching, suggesting that classroom teaching and assessment were not congruent with each other or with the course objectives. Thematic analyses of guided interviews with nine teachers provided insights into their grammar-oriented teaching practice. Along with washback, a lack of confidence in assessing students' speaking due to their anxiety as non-native English speakers emerged as an influential factor that hindered teachers from implementing the course objectives. Moreover, the results suggest that the high-stakes exams tend to be more influential than the educational policy in this context, especially in academically-oriented classrooms in which the majority of the students were hoping to go to university (as opposed to vocational-oriented classrooms). Finally, the thesis discusses the need for change in the high-stakes exam system as well as the urgent necessity of teacher training on assessment specifically designed for non-native language teachers, which would contribute to the improvement of EFL pedagogy.

Résumé

Des recherches ont démontré que l'effet de retour (l'impact des examens déterminants sur la nature et le déroulement de l'activité pédagogique dans les classes) était souvent partiellement attribuable aux enseignants (Cheng, 1997, 2000; Turner, 2001, 2006, 2009). L'idée comme quoi, dans les lycées japonais, à cause de l'influence fondamentale des examens d'admission universitaires, les cours axés vers la conversation, dits de communication orale (CO), ne s'articulent pas autour d'activités d'apprentissage interactives, mais plutôt autour de l'acquisition de notions grammaticales, est prépondérante dans la littérature sur le sujet. Cette thèse fait appel à des méthodes mixtes pour décrire l'état actuel des cours de CO dans le contexte de la politique nationale d'éducation japonaise, en adoptant la perspective des enseignants d'anglais comme langue seconde (ALS). Des données quantitatives récoltées à l'aide d'un sondage auprès d'enseignants (N=87), des données qualitatives fondées sur des observations dans les salles de classe ainsi qu'une analyse d'examens semestriels ont révélé que l'effet de retour était plus apparent dans les méthodes d'évaluation utilisées par les enseignants que dans l'enseignement prodigué, laissant supposer la présence d'incongruités entre les objectifs des cours, l'enseignement qui en découle et l'évaluation qui y est lié. Une analyse thématique d'entretiens dirigés avec neuf enseignants a permis de mieux comprendre leur méthode d'enseignement axée sur la grammaire. De pair avec l'effet de retour, leur manque de confiance quant à leur capacité à évaluer le discours oral de leurs élèves causée par une langue maternelle autre que l'anglais est ressorti comme un facteur significatif les empêchant d'accomplir les objectifs du cours. De plus, les résultats de l'analyse indiquent que dans un tel contexte, les examens déterminants avaient une influence plus prononcée que la politique éducative, surtout dans les programmes pré-universitaires (par opposition aux programmes professionnels). Cette thèse traite finalement du besoin de changer le système d'examens actuel et de la nécessité d'offrir aux enseignants dont la langue première n'est pas celle qu'ils enseignent une formation sur les méthodes d'évaluation adaptée à leurs besoins, ce qui contribuerait à l'amélioration de la pédagogie de l'enseignement de l'ALS.

Acknowledgements

I would like to express my profound gratitude to my supervisor, Dr. Carolyn. E. Turner, for her guidance and constructive feedback throughout all the various stages of this thesis. I would like to thank the second reader, Dr. Mela Sarkar, for her encouragement and mentorship. I am also grateful to the board of education and the EFL teachers in Japan, whose agreement to participate in the study enabled this research to take place. Additionally, I am truly appreciative towards the five teachers in the classroom observation segment during the most hectic time of the year.

Throughout the experience of my thesis writing journey, I met great people in Canada and have been encouraged by them. Although I would like to express my hearty thanks to all of my friends, I am sure it will be longer than this thesis. Thanks to my friends, my life in this far-away country has been enjoyable like the Japanese saying "よく遊びよく学べ" (play well, study hard). Alain and Luc, thank you for the great French translation, merci! Martin, thank you for your constant encouragement along the way.

Lastly, I extend my most sincere gratitude to my family for their unwavering support throughout my graduate studies and the process of writing this thesis. Particularly my mother, Yuriko, she is the only one who believes in me all the time, even when I am in the deepest adversity. Without her emotional support from Japan, I could not have come so far.

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Acronyms

AI: Academically-inclined; LAI: Less Academically-inclined

ALT: Assistant language teacher

CBA: Classroom-based assessment

CLT: Communicative language teaching

CLTT: Communicative language teaching and testing

CT: the Center Test

EFL: English as a foreign language; ESL: English as a second language

L1: First language; L2: Second language

MEXT: Ministry of Education, Culture, Sports, Science and Technology

MMR: Mixed method research

NS: Native speaker; NNS: Non-native speaker (of a language)

OC: Oral Communication

SLA: Second language acquisition

Chapter 1: Introduction

1.1 Background of the Context of the Research

Research has suggested that *washback*—the influence of testing on teaching and learning—is a multidimensional phenomenon. Alderson and Wall (1996) defined it as "the power of tests to affect what goes on in the classroom, the educational system, and society as a whole" (p. 236). Similarly, Bachman and Palmer (1996) noted that washback effects may appear in different forms depending on the contextual variables of the society in which the test is used. Shohamy, Donisa-Schmidt, and Ferman (1996) argued that if the stakes of a test are high in society, its influence over the stake-holders can also be strong.

In Japan, it is often argued that due to the tremendous significance of the university entrance exams in the society schools are likely to implement hidden curricula. Such uninvited impact from high-stakes exams is called *negative washback*. In terms of the effect on language education, there is the intense pressure on the shoulders of English as a Foreign Language (EFL) teachers to equip their students with the necessary skills to pass the high-stakes exams, which generally do not require speaking skills. Under such a circumstance, it is often argued that many EFL teachers retain the conventional grammar-translation methodology (Gorsuch, 2000, 2001; Kikuchi & Browne, 2009; Ressor, 2002; Taguchi, 2005; Watanabe, 1996, 2000).

In order to ameliorate this situation, over the last two decades, *the Ministry of Education and Science, Sports, and Culture* (MEXT) has been persistently promoting the use of Communicative Language Teaching (CLT) in the English classroom by revising the nationwide curricula, namely, *the Course of Study*. Accordingly, speaking-focused courses were introduced to the curricula in 1989. In 2003, MEXT announced an educational policy that aimed to produce

Japanese citizens with communicative competency in English to cope with globalization, which emphasized the importance of CLT to EFL teachers.

The CLT approach values the communication of meaning in interaction rather than focusing only on grammatical analyses (Lightbown & Spada, 2006, p. 196). In the same vein, as Bachman and Palmer (1996) propose, teachers should test students' communicative language skills of the target language in accordance with CLT. Owing to various social and cultural barriers, however, the tenet seems to be largely ignored in high school EFL classrooms in Japan (Gorsuch, 2000, 2001; Nishino & Watanabe, 2008; Taguchi, 2005). For reasons of the pedagogical predicament, teachers have asserted that CLT could not prepare their students for university entrance exams; thus, their teaching focus remains on developing students' grammar-translation skills. Moreover, Akiyama's (2004) and Taguchi's (2005) studies revealed that classroom testing, designed to be synchronized with teaching, was also influenced by the washback phenomenon; that is, teachers neither taught nor tested students' speaking skills even in speaking-focused courses.

Research has found that the occurrence of washback effects is often partly attributed to teachers (Alderson & Hamp Lyons, 1998; Cheng, 1997, 2000; Shohamy, 1994, 2004; Turner, 2001, 2006, 2009; Wang, 2010). Turner (2001, 2006, 2009) has claimed that degrees and kinds of washback occur through various contributing factors, and that these degrees and kinds are highly contingent upon teachers' perspectives and stances. Furthermore, researchers have posited that, to understand the complex nature of washback, teacher characteristics (e.g., personalities, educational backgrounds, working environments, and training experiences) should be taken into consideration. In the Japanese context, although some research has discussed the importance of the teacher factor in the occurrence of washback (Gorsuch, 2000, 2001; Hiramatsu, 2004;

Watanabe, 1996, 2000), little research has sought information from EFL teachers' perspectives on both classroom teaching and testing in speaking within the scope of negative washback.

In short, recent studies on washback suggest that teachers' influence over test implementation should not be ignored in order to explore the complex phenomena. As presented above, while previous studies in Japan confirmed the existence of negative washback from the deep-rooted significance of the university entrance examinations, few studies have closely investigated teachers' perceptions of negative washback effects on speaking. To provide an understanding of the stagnant situation where speaking skills are largely ignored in EFL classrooms, in spite of the constant promotion of CLT through the national educational policies, it is of importance to closely scrutinize the speaking-focused courses, especially teachers' attitudes and perceptions of the washback effect on their own testing practices.

1.2 Statement of the Research Problem

This thesis addresses problems in speaking-focused courses of English in Japanese high schools in relation to negative washback. In this context, teachers of English speaking courses have been facing a dilemma between the educational policy and the university entrance exams.

As noted earlier, in Japanese society, university entrance exams are considered as high-stakes exams. In particular, an annual standardized exam for university admission called "*the Center Test*," has significant importance in the test-driven Japanese society. Researchers have claimed that those high-stakes exams, which generally focused on testing receptive skills of English and ignored productive skills, were inducing negative washback in EFL classrooms (Hiramatsu, 2004; O'Donnell, 2005). It is a social norm that teachers in academically-inclined schools focus on test-taking strategies to prepare their students for the ultimate goal (Guest, 2008). Under the strong influence of university entrance exams, teaching methods of EFL

classrooms in Japanese high schools tend to consist of conventional, teacher-centered, and grammar-translation instruction.

On the other hand, the government has made attempts to promote the use of the CLT approach to change the conventional EFL teaching environment. Speaking-focused courses were introduced into the official curriculum of high schools in 1986 and implemented nation-wide from 1994. One point that should be noted here is that the national curriculum guidelines clearly require teachers to employ the CLT approach in classrooms; however, the guidelines do not specify that assessment should be coupled with teaching. What is more, in December 2008, the government announced that the forthcoming Course of Study, in effect from 2013, will pressure teachers into using English as a medium of instruction in English classrooms at the high school level. Such lack of coherence between the policy and testing system is causing a chaotic situation in the Japanese education system.

After two decades of the implementation of the speaking-focused courses, what needs to be scrutinized is EFL teachers' professionalism in classroom teaching and testing in relation to washback. Lastly, because the present study was conducted in the wake of the drastic announcement of the new educational policy as described above (i.e., the use of English as the medium of instruction), this investigation will differ from similar research conducted in the past.

1.3 Objectives of the Study and the Research Questions

The purpose of this study is, therefore, to examine the classroom teaching-testing congruence of the speaking-focused courses from EFL teachers' perspectives under the circumstances where the educational policy and the high-stakes exam system do not match. In particular, one of the central goals of this research is to juxtapose classroom teaching and testing. As noted, most of the previous studies on washback in Japan dealt with one aspect of classroom

practice, mostly on teaching, but little research has been done on testing/assessment. In other words, although the discussions of the test influence on teaching and learning largely and vaguely embrace classroom-based assessment (CBA) as an element of what goes on in classrooms, the importance of CBA itself, separated from the comprehensive term of teaching, is likely to be less highlighted in the washback literature. As opposed to the washback effect '*of testing*' (the influence of a high-stakes test), the effect '*on testing*' (the influence on classroom testing/assessment), is still not a fully examined aspect of language assessment in this context. Hence, this study was conducted with the aim of providing a more holistic outlook on the issue by contrasting what teachers teach and how they test in classrooms.

Moreover, considering the importance of the teacher factor in washback, the second goal of the study is to document how EFL teachers in public high schools in Japan perceive their profession under such circumstances. One particular focus of teacher-related variables in this study is "working environments." Teachers in academically-inclined schools and ones in less academically-inclined schools may have different attitudes and perspectives towards their classroom practices. This aspect is seldom examined in the literature; therefore, this study aims to build on prior research and to contribute to a better understanding of the entire issue of washback.

To be able to meet the broad aim set out above, the following research questions are posed to guide this study:

1) How are teaching and testing in speaking-focused courses of English related to each other when the educational policy and the high-stakes exam are not in alignment?

2) What are the nature and the scope of washback effects from the university entrance exams on teacher perceptions in the speaking-focused courses?

To achieve the abovementioned goals, I visited a local prefecture in Japan to collect data through a survey, classroom observations, term exams, and interviews. The participants were Japanese EFL teachers working in public high schools. Due to my teaching experience in the same context, I was able to access the participants with help from the Board of Education. My personal and professional experience—as a learner and as a teacher of EFL—guided and motivated me to carry out this research.

1.4 Structure of the Thesis

This chapter provides an introduction to the context of the study and presents the aim and research questions that the thesis attempts to answer. In Chapter 2, a review of the literature discusses washback in relation to language testing. It starts with an overview of the theoretical advances in the definition of washback before reviewing the issues within the phenomenon. After discussing the definitions of washback, it looks at studies that have identified a reoccurring theme in washback studies, which is the 'teacher factor'. Then, the second section of Chapter 2 discusses washback studies conducted in Japan to help conceptualize the pedagogical issue that this study investigated. Within the scope of washback, it touches on the two contradicting factors—university entrance exams and the national educational policy—that make the target context unique and complex. Then, a brief theoretical and historical view of communicative language teaching and testing is presented in order to highlight the issue of speaking-focused courses of English in Japanese high schools. Lastly, it reviews some key studies on negative washback effects on the speaking-focused courses, which is the context of this study.

Chapter 3 describes the methodology and research design of this study. Beginning with the rationale for the methods used to conduct the research, detailed information on the

participants, the instruments, the data collection procedures, and the data analysis methods are presented.

The results section, Chapter 4, discusses the most significant data that are relevant to the research questions. First of all, in relation to washback, the updated reality of classroom teaching practices in the speaking-focused courses is presented, and that of testing practices follows. In relation to the results of the classroom practices, teacher perspectives of the influences on the courses are addressed. A short summary and reflection follow directly after the chapter in which the the research questions are addressed.

Chapter 5 discusses the main results in relation to certain background variables and related research in the area. The discussion broadly follows the research questions and builds on the analyses of the findings described in Chapter 4.

Chapter 6 focuses on the conclusions drawn from this research as well as implications and suggestions. Limitations of the study are also presented in this section. The chapter interprets the results, links them with the research literature, draws conclusions, and examines implications for policy makers, EFL language education, and future research.

1.5 List of Definitions of Terms

Below I introduce a list of definitions of terms and concepts that will be used frequently in this thesis. These definitions are provided in order to guide readers and should be interpreted as such within this specific context.

Assessment / Testing: In this thesis, the two terms are used interchangeably to refer to the process of documenting a learner's knowledge and skills of English.

Classroom practices: This term is used to refer to what teachers do in classrooms to facilitate students' language learning, which includes not only teaching but also assessment/testing organized and administered by the teachers at the classroom level.

The Center Test (CT): This test is a standardized test developed by the National Center for University Entrance Examinations (NCUEE), and used for admission to a large number of Japanese tertiary institutions such as all national and public and many private universities.

High-stakes exams: The term is used to refer to tests on which test-takers' career or study plans hinge. The tests are likely to be influential on the teaching and learning behaviours of those involved in the tests. Thus, even a minor change in the test can cause strong washback effects for the stake-holders (Shohamy, 1996).

Low-stakes exams: As opposed to high-stakes exams, low-stakes exams refer to tests that do not entail serious consequences for the test-takers' future plans. School-based and classroom-based tests/assessment falls into this category in this study.

OC: "Oral communication" (OC) is a title for English courses that mainly focus on developing learners' aural/oral skills. The courses were introduced in 1989 in Japan to promote the use of CLT in classrooms.

University entrance exams: In this thesis, university entrance exams mean external EFL tests conduced by individual universities. While the Center Test is one type of university entrance exams, this term encompasses a wide range of entrance exams for university admission.

Washback: This term is used to refer to the influence on teaching and learning generated by a test (usually a high-stakes test). The influence can be either positive or negative (Bailey, 1996). Washback is further defined in the next chapter.

1.6 Summary

Despite the strong political movement to promote the use of CLT in EFL classrooms in Japan, university entrance exams remain in a traditional grammar-oriented form. In such an incongruent educational system, EFL classroom teachers have been in a quandary for decades and tend to choose teaching towards the tests. To date, most of the research investigating this issue in relation to negative washback has focused on classroom teaching, and the influence of the university entrance exams on classroom testing/assessment has likely been beyond its scope. Therefore, this study was conducted with the aim of documenting the updated reality of speaking-focused courses by examining both teaching and testing of the courses. Moreover, scrutiny of teacher perceptions on influential factors on their classroom practices—teaching and testing—provides deeper insights into their professionalism under the incoherent circumstances.

This chapter outlined the background of the research issue and the rationale of the study. In addition, it described the motivation for focusing on Japanese high school EFL education and justified research questions that the thesis attempts to answer. Finally, the chapter provided an overview of the main components of the thesis as well as definition of key terms used in this thesis.

Chapter 2: Review of the Literature

2.1 Introduction

This chapter focuses on previous studies and is comprised of three sections to illustrate the concept of washback and the specific context in which I conducted my research. The first section provides an overview of the theoretical framework of washback to gain insight into the complex nature of the phenomenon. Following the historical perspective of washback in second language (L2) and general education, it looks at the definitions of washback and related concepts to construct the significance of this issue to be further investigated. Then attention shifts to the importance of the teacher factor as a reoccurring theme in washback studies. The second section focuses on washback research in Japan. With a brief introduction to the background of the EFL education and the national high-stakes exam, this section introduces studies on washback specifically conducted in Japanese contexts. Lastly, the third section is devoted to the issue of the speaking-focused courses at the high school level in Japan, which is the target context of this study. Due to the lack of a speaking component in most of the university entrance exams, the dearth of correspondence across teaching, testing, and curriculum of the courses have been evoking heated discussions among washback researchers. This section also illustrates the concept of Communicative Language Teaching and Testing (CLTT) to highlight the issue of the curriculum-testing incongruence in this context.

2.2 The History of Washback Studies and the Definitions

To date, a burgeoning number of studies have reported the phenomenon of washback both in the field of general education and in language education. While the explanations of the term 'washback' can differ over time and according to the context of research, the term commonly refers to the influence of testing on teaching and learning (Cheng, 2000). Additionally,

a variety of terms have been used to refer to this notion in addition to washback (Alderson & Wall, 1993; Buck, 1988), such as backwash (Hughes, 1989, 1993) and test-impact (Bachman & Palmer, 1996; Baker, 1991). For the sake of consistency, and given that most of modern language testing researchers currently use the term 'washback', I will follow the mainstream in this thesis and use this term to refer to the central phenomenon under study.

In the 1950s and 1960s, researchers in general education started to point out that exams could have some influence over teachers, learners, and other stake-holders. In the literature, the term "exams/tests" largely refers to high-stakes ones that are external to the classroom. Vernon (1956), for example, claimed that teachers had a tendency to ignore content that would not appear in the exam, by distorting the curriculum. Davies (1968) argued that tests and test materials were likely to be used as teaching devices to narrow down the teaching content. In the earlier era of language testing research, the late 1980s, researchers started applying the concept of washback to language testing contexts. Swain (1985) stated that language teachers would teach content to their students in accordance with the test. Buck (1988, cited in Bailey, 1999), knowing the test-driven society in Japan, that Japanese university entrance exams apparently had a strong impact on teachers and students, and stated that the influence made them "tailor their classroom activities to the demands of the test (p. 17). In his book for language teachers, Hughes (1989) asserted that it is of great importance for teachers as testers to be aware of the effect of testing on learning outcomes.

Up to the 1990s, wherein the apparent existence of washback was widely acknowledged by researchers, there was little empirical data to define the concept. In 1993, Wall and Alderson conducted the first empirical study on washback in Sri-Lankan secondary schools in order to investigate the effects of a new English examination on classroom teaching. They concluded that

the tests had impact on "what teachers teach but not on how they teach" (p. 68), indicating the complex relationship between testing and teaching. Lam (1994) also conducted empirical research in Hong Kong and observed that a revision of a high-stakes English exam could bring about a positive impact on teachers and learners. In terms of test stakes, Shohamy et al. (1996) pointed out that low-stakes exams are unlikely to be as influential as high-stakes exams. From questionnaire data collected in Israel, they found that the low-stakes exam of Arabic as a second language had little impact on students as opposed to the high-stakes EFL exam. In her study on the possible washback effect from the high-stakes English exam in Hong Kong secondary schools, Cheng (1997) reported that although a revision of the high-stakes exam could generate intended washback effects on language teaching and learning, "unintended and accidental side-effects could occur" (p. 68). This study revealed the "complexity of multiple influences" (Cheng, 2005, p. 25). Research on washback then rapidly expanded its scope from the simple relationship between testing, teaching and learning to the more complex nature of the test influence. In this sense, Bachman and Palmer (1996) have argued that test impact can reach the stake-holders at the micro level (individual teachers and students) and at the macro level (society and educational systems).

In relation to the expansion of the definitions, Hughes (1993) has provided a useful framework to clarify how a test may affect teaching and learning. In his framework, stake-holders are called "*participants*" (e.g., teachers, students, material developers, publishers, and administrators). Then, the participants perform "*processes*," which indicate any sorts of actions and decisions to contribute to the process of learning (e.g., changes in teaching or learning practices, curriculum design, and instructional material creation). Then, the processes yield the "*products*", which is the outcome of the learning experiences. According to Hughes, ideally, a

test should contribute to the improvement of the product through the processes. This framework

was adapted by Bailey (1996) who created a model of washback to illustrate the inter-

relationship of these mechanisms (see Figure 1).
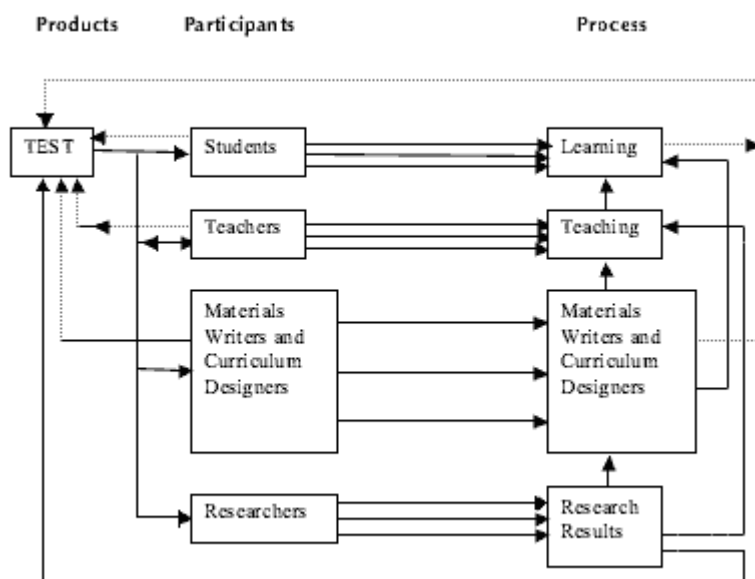
Figure 1 A basic model of washback (Bailey, 1996)

In this model, Bailey presented the participants-processes-products trichotomy in a slightly

different manner from that Hughes. She hypothesized that the attitudes and the perceptions that

the participants hold towards the test are the processes; for example, some teachers (participants)

consider that the test results can change their students' future plans (process) and change their

usual teaching practices (products) for the sake of better learning outcomes (the ultimate

products). This model is very useful, shedding light on the crucial role of the participants'

attitudes and perceptions in the nature of washback. In other words, how the participants

perceive the test may be attributed to whether it can generate beneficial washback effects on

learning outcomes or not. More recent studies on washback have focused on diverse factors

involved in the mechanisms of washback. Turner (2001) pointed out that the mechanisms of washback could vary depending on the socio-cultural, socio-political, contextual factors, and the participants involved in the test implementation process. In the following section, this review explores studies that investigated different mechanisms of the washback phenomenon in relation to the consequences of the test.

2.2.1 Negative and positive washback.

In general, washback can be either positive or negative, depending on whether the test improves or deteriorates the quality of the learning outcomes (Bailey, 1996). Oller, back in 1979, stated that a good test possesses characteristics of reliability, validity, practicality, and instructional value. The last characteristic indicates that an ideal test should contribute to the improvement of instruction, which is close to the notion of "positive washback." Alderson and Wall (1993), noting that washback can have harmful effects, coined the term "negative washback"; namely, it is the unsought influence of a test on teaching and learning, such as unreasonable reduction of subjects and activities. Messick (1996) also described it as "the extent to which the introduction and use of a test influences language teachers and learners to do things they would not otherwise do that promote or inhibit language learning" (p, 241). In a literature review article, Pan (2009) summarized positive and negative washback as follows (see Table 1 & 2).

Table 1 *Summary of positive washback* (Pan, 2009, p. 261)

Positive Washback	
Classroom settings	1. Tests induce teachers to cover their subjects more thoroughly, making them complete their syllabi within the prescribed time limits. 2. Tests motivate students to work harder to have a sense of accomplishment and thus enhance learning. 3. Good tests can be utilized and designed as beneficial teaching-learning activities so as to encourage positive teaching-learning processes.
Educational/societal system	Decision makers use the authority power of high-stakes testing to achieve the goals of teaching and learning, such as the introduction of new textbooks and new curricula.

Table 2 *Summary of negative washback* (Pan, 2009, p. 261)

Negative Washback	
Classroom settings	1. Tests encourage teachers to narrow the curriculum and lose instructional time, leading to "teaching to the test."
	2. Tests bring anxiety both to teachers and students and distort their performance.
	3. Students may not be able to learn real-life knowledge, but instead learn discrete points of knowledge that are tested.
	4. Cramming will lead students to have a negative [perception] toward tests and accordingly alter their learning motivation.
Educational/societal system	Decision makers overwhelmingly use tests to promote their political agendas and to seize influence and control of educational systems.

Regarding negative washback, some researchers have provided empirical data to make a case for this issue. A survey study conducted by Alderson et al. (1990) in Canadian secondary schools documented that teachers narrowed down the topics that were likely to be included in the final exam, and that the learning focus was more on rote memorization than critical thinking. Shohamy et al. (1996) reported negative washback in a study that investigated the effects of high-stakes EFL exams in Israel. In the study, the teachers quickly shifted their teaching focus from oral proficiency to writing skills once the oral test had been finished to prepare their students for the following paper-based exam. The study of Wall and Horák (2006) revealed that

teachers were likely to emphasise memorization and test-taking techniques as preparation methods in Central and Eastern Europe. They investigated the impact of significant changes in the TOEFL exam on the participants of the test preparation courses. The new speaking components in the exam were included so as to encourage communication-oriented teaching methods; however, it was found that the teachers' focus was still on pattern drills that mirrored the new test tasks. In their conclusion statement, sufficient teacher training was stressed out as a key factor of a successful test implementation. In this regard, Watanabe (2000) also reported the crucial role of teachers in negative washback. This aspect will be discussed further in the following section.

Some studies, in contrast, have documented desirable learning outcomes generated by exams. Positive washback is in close relation to the concept of "measurement-driven instruction"; that is, "the notion that tests should drive learning" (Shohamy, 1993, p. 4). As one example of positive washback, in the Canadian context, Turner (2001) found that teacher involvement in the test design process could contribute to the promotion of positive washback by itself. She reported that teachers involved in the development of rating scales displayed more positive attitudes towards the exam, which were reflected in the teaching practices consequentially. Building on Turner's (2001) findings, Saif (2006) developed a similar study to investigate whether the participants' involvement in the test development process would result in positive washback in a Canadian university. The results suggested a positive relationship between the test and teaching and learning outcomes. In their study on classroom-based assessment of oral proficiency, Muñoz and Álvarez (2010) also documented positive washback in some of the areas examined. They claimed that constant guidance and support for teachers was vital to create positive washback.

2.2.2 The mismatch between the curriculum and the tests.

As discussed above, washback has been conceptualized as a complex phenomenon, which can be positive or negative, depending on the factors involved in the test. In relation to the bipolar characteristics of washback, it would be helpful to further touch on the effects of washback on curriculum. In terms of negative washback, it has often been noted that a mismatch between the curriculum and the high-stakes exam's content can lead to harmful influences on classroom teaching and learning. When a high-stakes exam is not designed to appropriately reflect the curriculum goals, the teachers whose students will take the exam may abandon the curriculum and accommodate their teaching to the exam content. As noted above, studies on negative washback reported the narrowing of the curriculum to specific areas that would be tested (Alderson et al, 1990; Lam, 1994; Shohamy et al, 1996; Wall & Horák, 2006).

On the other hand, to generate positive washback, "the ideal situation in an education system is when the curriculum, teaching and testing are synchronized" (Turner, 2009, p. 105). In this sense, the aforementioned studies conducted by Turner (2001), and Saif (2006) showed that when the curriculum and testing were in alignment, teaching tended to be synchronized with them.

The importance of the curriculum-teaching-testing congruence is a key theme of the research context in this study; therefore, it will be further discussed in the following section. Prior to that, it is important to explore another reoccurring key theme in washback literature, which is the teacher factor. Amongst numerous factors involved in the complex phenomenon of washback, one of the most salient findings emerging from studies is that teachers play a pivotal role when washback occurs.

2.2.3 The teacher factor in washback.

As presented in Bailey's model (see Figure 1), teachers as participants are situated right under students, which indicates they have direct influence on students' learning outcomes. Thus, some teachers' negative attitudes and feelings towards high-stakes exams can influence the ultimate product, which is the students' learning outcome. In this regard, Cheng's (1997) study is worthy of note. She investigated whether a revision of the Hong Kong Certificate of Education Exam in English, whose stakes were high in that context, would bring intended positive washback to the test-takers. At odds with the expectations of the authorities, teachers showed little change in their teaching practices. In fact, they changed the content of teaching but did not adapt a new instructional method by maintaining the traditional teacher-centered environment. The important implication of this study is the difficulty of inducing positive washback and the significance of teachers' role in the mechanism. More recently, Wang (2010) conducted a study, also focusing on the teacher factor in a Chinese university context, and documented that while testing (a high-stakes EFL exam in this case) had limited washback effects on classroom practices, teachers' individual characteristics such as pedagogical beliefs, instructional knowledge, and prior experiences served as a framework for their teaching practices. She concluded that even though the external force of tests might bring about washback on teaching and learning, the internal force of classroom teachers' attitudes and perceptions surpassed it and determined the degree of the test influences.

We see, therefore, when it comes to generating positive washback, manipulating the teacher factor for such purpose can prove challenging. Lam (1994) noted that teacher characteristics varied and so did their attitudes and perceptions of the test. This appeared to influence their teaching practices differently. He observed that some teachers were more

innovative in creating new materials on their own to cater to the revised English exam whereas others were not. Turner (2001, 2006, 2009) has claimed that degrees and kinds of washback occur through various contributing factors, and that these degrees and kinds appear contingent upon teachers' perspectives and stances. Bailey's (1996) call that the validity of exams is a crucial factor contributing to positive washback remains a key factor of positive washback. Along with that, based on the findings in the literature, it is of importance to consider teacher characteristics (e.g., attitudes and perceptions towards high-stakes exams, personalities, educational backgrounds, working environments, training experiences) to discuss the phenomenon (Turner, 2006, 2009).

However, despite a growing body of literature that highlights the pivotal role of teachers in washback, there is still a dearth of conceptual evidence on how a high-stakes exam may influence teacher perception of classroom assessment practices. Researchers have claimed that teachers tend to play a powerful role in inducing washback by reflecting their perceptions of the exam in their teaching practices. It is necessary to document their professionalism as testers at the micro level in a mechanism of negative washback to provide an understanding of the role of the teacher factor.

2.3 Research on Washback in the Japanese Context

In this section attention shifts to the target context of this study, which is the EFL education system in Japan. Within the scope of washback, this section provides the overall background information of the specific setting. Firstly, an overview of the significance of the university entrance examination in Japanese society is provided based on the literature that examined negative washback effects generated by high-stakes exams. Secondly, this section discusses studies on the incongruence of the national educational policy of EFL pedagogy and

actual classroom teaching. Next, given the focus of this research on problems regarding the

teacher factor embedded in this specific context, the last part of the section explores Japanese

EFL teachers' attitudes and perceptions concerning washback.

2.3.1 The significance of the university entrance examination.

As noted earlier in Chapter 1, it is widely known that Japan is a highly test-driven society.

As early as Buck (1988), this situation was recognized and described as follows:

> "Japan is a country in which the entrance examination reigns supreme. It is almost
>
> impossible to overstate the influence of these examinations on both the educational
>
> system as a whole, and the day-to-day content of classroom teaching. Their importance in
>
> the lives of young people is such that almost all future social and economic advancement
>
> is dependent on the results of these entrance examinations" (p. 16).

Although this statement was made over 20 years ago, it is applicable to the modern setting. To

date, many researchers have cast doubt on the validity of university entrance exams and

questioned their consistency with the goals of the curriculum. Brown and Yamashita's (1995)

study confirmed how inconsonant the Japanese testing system was with the national curriculum,

namely, "*the Course of Study*," which will be further discussed in the following section.

Examining entrance exams at a prestigious private university, they concluded those exams were

merely designed to measure test-takers' language analytic techniques; therefore, the teacher-

centered grammar-oriented methodology was overwhelmingly dominant (Brown & Yamashita,

1995). Based on this, Kikuchi (2006) replicated their study ten years later to monitor changes in

such exams in the same context. Although some minor changes were found, he reported that "the

types of items, their variety, and the skills measured did not look substantially different"

(Kikuchi, 2006, p. 77). Researchers have considered the exams as a cause of such a chasm

between the national guidelines and actual classroom practices. As an updated reference, Kikuchi and Browne (2009) explicated the reality of English courses from the perspective of students just enrolled in universities. The student survey revealed that their EFL teachers in high schools did not effectively implement the Course of Study guidelines. They concluded that the powerful influence of the exams remained unchanged in the classroom.

2.3.1.1 *The Center Test.*

It should be noted that there exists one particular university entrance exam, namely *The Center Test*, which is extremely high-stakes in Japan. The Center Test is a standardized annual test developed by the National Center for University Entrance Examinations (NCUEE), and used for admission to all national and public and many private universities. Admission to national and public universities is based mainly on the results of two entrance examinations. In January, students have to take the Center Test and wait for the result until February. If the score is high enough to satisfy the criteria of a university she/he wants to enter, the student is able to take an additional exam that is conducted by the individual university in February or March. The Center Test covers Japanese language and literature, social science, mathematics, science and foreign language. Students have to take exams in assigned subjects, depending on the academic field they wish to study at the university.

In 2011, 99.84% of test-takers chose to be tested in English as a foreign language, whereas Chinese, Korean, German, and French were chosen 0.06%, 0.03%, 0.03%, and 0.03% respectively (NCUEE, 2011). This is because most of the tertiary institutions require the score of English language along with other subjects. This statistical data indicates the significance of English in the field of foreign language teaching in Japan. Inferring negative washback effects, Guest (2008) described a substantial impact on teaching and learning as follows:

"Thousands of *juku* (cram schools) and almost all academically-inclined high schools

base their final year curriculum around achieving success on this exam, hence its

allegedly profound effect on secondary education" (p. 16).

In the format of multiple-choice questions, the Center Test of English lacks both writing and

speaking tasks. A listening component was finally added in 2006. When it comes to productive

skills, however, it seems unfeasible to ensure the appropriate evaluation of such skills (i.e.,

speaking and writing skills) due to the sheer number of test-takers (i.e., over 500,000 students).

Moreover, like the Center Test, other university entrance exams conducted at individual

universities generally do not include a speaking component. Therefore, EFL teachers in high

schools tend to focus on developing students' receptive skills at the cost of productive ones.

Despite the importance of this particular exam, there is limited research on the influence of this

exam on classroom practices.

In short, the strong influences of the university entrance exam, in particular the Center

Test, do not seem to be waning at present, and, in particular, in Japanese society. As discussed in

the literature, as the validity of the Center Test of English language is questionable (Brown

&Yamashita, 1995; Guest, 2008; Kikuchi, 2006), it generates negative washback (Messick,

1996). In an attempt to ameliorate the stagnant situation, the government has been promoting the

use of CLT in classrooms by revising the Course of Study for over the last twenty years.

However, as a result of constant pressure from the education authorities, the issue of the

curriculum-testing incongruence is becoming increasingly more complex in this context.

2.3.2 The national curriculum.

Educational policies are created in a top-down manner in Japan. Following the MEXT's

revision of the policy, each school organizes its own curricula within the framework of the

Course of Study. In the current national guidelines for English courses, the strong emphasis is on the importance of developing students' ability to use English for the purposes of everyday communication (Gorsuch, 2000). Furthermore, in 2002, the Action Plan to Cultivate "Japanese with English Abilities" was announced in order to urge EFL teachers to be aware that their mission was to equip students with "communication skills in English" (MEXT, 2002).

As noted earlier, without fundamental changes in the university entrance examination system, the goals of the educational policies do not seem to be achievable. Taking a critical position on the issue of the pedagogical discrepancy between the goals of educational policies and the reality of grammar-focused classroom teaching, Reesor (2002) stated that the Japanese commitment to foster more competent English language speakers was "half-hearted at best and downright false at worst" (p. 42).

Under such a double-bind circumstance between the educational policy and university entrance exam, it is necessary to investigate how Japanese EFL teachers position themselves as professionals in order to understand the negative washback phenomenon in this context. Therefore, next, I explore the teacher factor reported in the literature.

2.3.3 The teacher factor in the Japanese context.

In line with the view held by washback researchers in other contexts (Alderson & Hamp Lyons, 1998; Cheng, 1997; Shohamy, 1994, 2004; Turner, 2001, 2006, 2009), in Japan, Watanabe (1996, 2000) also found that teachers played a pivotal role in determining whether washback occurred. In his 1996 study, Watanabe reported that teachers narrowed teaching contents, excluding listening and writing, to prioritize other skills in which it was likely to be easy to gain high scores. In his 2000 study, it was pointed out that teachers' instruction focus was on test taking techniques but not on listening and speaking skills. Based on these findings,

Watanabe has claimed that teachers' psychological factors, such as educational background, personal beliefs, and teaching experience could generate negative washback. The data of the two studies, however, were collected from teachers working in a private cram school (*yobiko*) where students go with an expectation of learning specific skills to pass university entrance exams. It should be noted that teachers in such cram schools are not necessarily to have teaching credentials/certifications unlike EFL teachers in public high schools. Therefore, it is questionable to apply the same theory to teachers working in public schools within the framework of the Course of Study, and it is thus important to document if those EFL teachers display different professionalism from the ones in cram schools.

In this sense, Gorsuch's (2000) survey of teachers' perceptions of the influences on their classroom practices helps provide clarity about the target context of the research carried out and described in this thesis. The sample size for the survey was 876, all high school EFL teachers, and the results indicated that the factor of university entrance exams displayed its significance in this context. Gorsuch identified several factors that seemed to intertwiningly affect teaching practices. The factors included university entrance exams, class size, textbooks, students' and teachers' English speaking ability, and expectations of students and parents. Amongst these factors, the centrality of university entrance exams was significant. Thus, these results indicate that the deep-rooted importance of the high-stakes exams in Japanese society forces teachers to retain the conventional grammar-translation methodology, since intense pressure is on the shoulders of EFL teachers to equip their students with the necessary skills to pass high-stakes tests.

Similarly, Sato (2002) conducted classroom observations in a public high school, and found that due to the long-lasting negative washback, teachers were unwilling to adopt new

teaching methods even after intensive CLT training. He reported that teachers' adherence to grammar-translation methods did not change, as they had to go back to the same teaching routine. Likewise, in her study on constraints on adopting CLT in high school classrooms, Cook (2009) described the teacher role in a school. In Japanese society, schools were expected to guarantee parents and students acceptance into universities, and as a result, teachers paid less attention to skills not required for the entrance exams. Additionally, she listed other influences on classroom teaching in this context, such as school academic levels, teachers' frequent transfers, and the schools' hidden role as service provider. Cook described the distinctive role of high school teachers in Japanese society the as follows:

"A strong underlying reason why teachers seem to be giving into the demands of their students, colleagues, and society in general is that paying parents (at the high school level) expect to be guaranteed that their children will succeed on entrance exams… In the end, teachers are forced to provide a specific service or face public criticism" (p. 112).

Those unique characteristics, which constitute the mechanism of negative washback in Japan, have not been fully explored yet. However, considering the fact that teacher-related factors (their beliefs, attitudes, educational level and experience, and personalities) can affect washback (Spratt, 2005), these characteristics need to be further investigated.

2.4 Speaking-focused Courses "Oral Communication"

In order to conduct my research on negative washback, I chose one particular setting which has been evoking heated discussions among scholars, that is, oral communication courses. The debate has been over the inability of the courses to attain the expected outcomes of students' linguistic development. This section first looks at the overview of the courses to define the research context. Additionally, to illustrate how negative washback emerged in the speaking-

focused courses, it defines the concept of CLTT for this research. Finally, it reviews related studies on negative washback effects on teaching and classroom testing practices.

In tandem with the abovementioned changes of national curriculum (see Section 2.3.2), speaking-focused courses—Oral Communication A, B, and C—were implemented starting in 1994 in high school curricula in order to ameliorate the previous situation where the target language was taught like Latin. In 2003, in accordance with the revision of the Course of Study, the speaking-focused courses were "improved" to put further emphasis on more authentic and practical English use. The courses were named Oral Communication I and II. For Oral Communication I, the course objectives are defined as:

> "To develop students' basic abilities to understand and convey information, ideas, etc. by listening to or speaking English, and to foster a positive attitude toward communication through dealing with everyday topics" (MEXT, 2003, section II, para.1).

For Oral Communication II, the course objectives are stated in this way:

> "To further develop students' abilities to organize, present and discuss information, ideas, etc. in English, and to foster a positive attitude toward communication through dealing with a wide variety of topics" (MEXT, 2003, section II, para.2).

In sequence, the two courses are intended to focus on aural and oral skills and are expected to be organized and developed within the framework of the specified course objectives. One important

oversight, however, is that the official guidelines lack any suggestions pertaining to assessment and how it should be integrated into classroom practice.

2.4.1 Communicative language teaching and testing.

As noted in Chapter 1, OC courses were created to promote the use of CLT, and should be taught in the framework of the CLT approach. Thus, it is beneficial to briefly introduce the concept of CLT before introducing literature that has identified problems in OC courses. Additionally, to discuss the importance of testing that complements students' language learning experience, the definition of Communicative Language Teaching and Testing (CLTT) is also provided in this section.

Supported by Swain's Output Hypothesis (1985) and Long's Interaction Hypothesis (1996), it is largely claimed that learners need the opportunity for *pushed output* through comprehensible interaction in order to enhance their language acquisition experience. Based on this theoretical foundation, CLT emphasizes the importance of interaction in the target language rather than focusing only on grammatical analyses (Lightbown & Spada, 2006). To define the nature of CLT, Brown (2000, p. 43) proposed a summarized overview by listing six characteristics as follows:

1) Classroom goals are focused on all of the components (grammatical, discourse, functional, sociolinguistic, and strategic) of communicative competence. Goals therefore must intertwine the organizational aspects of language with the pragmatic.

2) Language techniques are designed to engage learners in the pragmatic, authentic, functional use of language for meaningful purposes. Organizational language forms are

not the central focus, but rather aspects of language that enable the learner to accomplish those purposes.

3) Fluency and accuracy are seen as complementary principles underlying communicative techniques. At times fluency may have to take on more importance than accuracy in order to keep learners meaningfully engaged in language use.

4) Students in a communicative class ultimately have to use the language, productively and receptively, in unrehearsed contexts outside the classroom. Classroom tasks must therefore equip students with the skills necessary for communication in those contexts.

5) Students are given opportunities to focus on their own learning process through an understanding of their own styles of learning and through the development of appropriate strategies for autonomous learning.

6) The role of the teacher is that of facilitator and guide, not an all-knowing bestower of knowledge. Students are therefore encouraged to construct meaning through genuine linguistic interaction with others.

In line with the theoretical and empirical development of CLT, scholars' attention was also directed to testing/assessment that complemented this particular teaching approach. From the perspective that language learning processes should not be completed without assessment, Bachman and Palmer (1996) proposed that if teachers followed the policy and implemented CLT, assessment should be coupled with the instruction; in other words, teachers should test students' communicative language skills of the target language to complement their learning experiences. In his article on communicative language testing, Skehan (1990) contended that it was common

sense that "performance on a paper-and-pencil and multiple-choice grammar test only generalized hazardously to performance in real-life situations" (p. 120).

In the same vein, Canale and Swain (1980) claimed that by reason of a theoretical distinction existing between competence and performance, the learners had to be tested not only on their knowledge of language, but also on their ability to put it to use in a communicative situation. Based on the theoretical foundation of Communicative Language Teaching and Testing (CLTT), when a communicative approach to second language teaching is adopted, principles of syllabus design must include not only grammatical knowledge but also sociolinguistic and strategic competence (Canale & Swain, 1980). Bachman (1990) then applied and extended the framework of language competence created by Canale and Swain (1980). His model is comprised of three components: language competence, strategic competence, and psycho-physiological mechanisms. Bachman and Palmer (1996, 2010) have extended this model. These influential models of language competence provide useful guidelines for test developers and contribute to the field of language testing. Building on this theoretical foundation, Moon and Callahan (2001) state that performance assessment can promote test authenticity because it situates learners to perform in the target language, which simulates real-life experiences or problems. Therefore, teaching and testing/assessment should correspond in educational settings and hence, CLT should be accompanied by communicative testing. Both need to be in sync to help learners achieve curriculum aims.

2.4.2 The negative washback in the OC courses.

OC courses are supposed to be organized and taught based on the abovementioned theories. However, the literature indicates that the reality does not seem to answer the expectations of MEXT. That is, the continuous attempts to promote the use of CLT in Japan have

failed to be accepted at the classroom level. This part reviews some key studies that discussed problems arising from OC courses in relation to negative washback.

2.4.2.1 *Teaching practices*.

Hiramatsu's (2004) qualitative, and Kikuchi and Brown's (2009) quantitative, studies depicted similar scenarios in which the OC courses were not synchronized with the course objectives. Hiramatu revealed that EFL teachers avoided adopting CLT for the sake of test preparation through the teacher interviews, and Kikuchi and Brown documented a similar situation where the courses were not taught in accordance with the course objectives based on the students survey.

Amongst several studies in the literature that addressed the reality of the courses, Taguchi's (2005) study, which used combined methods of a survey and class observations, aptly describes characteristics of OC classes. The study revealed that the traditional form-based teaching method was still dominant in the speaking-focused courses, claiming "the national curriculum seems to remain at a formal level and does not achieve its functional ability" (p. 10). Regarding activities used in the classrooms, the environments observed by the researcher were highly teacher-centered and form-focused. Listening exercises and dialogue practices were found to be the most frequently employed activities, but meaning-focused activities for speaking were seldom used. In terms of instructional language use, Taguchi's survey revealed that 93% of the teacher participants answered using Japanese as a medium of instruction. She associated this result with negative washback, as the focus of the lessons was on grammar-translation skills.

With 268 student participants from two universities in Tokyo, Ichikawa (2006) conducted a survey study to examine students' perceptions of their OC class experiences. He reported that although there seemed to be major differences of course organization depending on each school,

largely the classes were constructed in the traditional manner to prepare them for university entrance exams. With reference to "communicative competence", only 22% of the participants agreed that OC courses helped them develop their communication skills in English. This fact made him conclude that the courses were not successful (p. 246). His qualitative data revealed that some students even felt demotivated because of their classes mismatching their needs and expectations towards the OC courses. That is, the courses were not focused on oral skills but on grammar to prepare the students for university entrance exams. He then proposed that EFL teachers should accept and perform their responsibility as professionals by using English to conduct lessons and, considering students' needs, to develop a more appropriate school curriculum.

2.4.2.2 *Testing practices*.

When it comes to language testing in Japan, research concerning classroom-based testing in relation to washback is scant in the literature. Very few studies to date have documented classroom testing in OC classes. In their qualitative research on the relationship between EFL teacher training experiences and student learning in a Japanese public high school, Sato and Takahashi (2003) reported that teachers who were in charge of OC classes expressed their doubt about the effectiveness of oral performance tests. They quote one teacher who was frustrated with the expediency of speaking tests: "Should we continue oral performance tests when they take too much time to prepare and the students don't improve?" (p. 330). Constrained with time and the pressure of preparing students for university entrance exams, the teacher participants could not be persuaded to adapt communication-oriented teaching and testing, but maintained the traditional grammar-translation methods.

Akiyama (2004) ventured to touch on the fact that teachers tended not to conduct speaking performance tests even after having speaking activities in the classroom, due to speaking skills not being required for the university entrance exams. In his argument, engrained cultural values in the school context were regarded as the reason for the gap between classroom teaching and testing. Although the definition of "speaking activity" may vary from teacher to teacher, the study questioned whether the curriculum, teaching and testing were congruent in this context. Instead of speaking tests, as Taguchi (2005) reported in her qualitative study, teachers tended to use paper-based tests in their classrooms as a summative assessment tool in OC classes. In general, in-class exams are designed by teachers themselves in each school to assess their students' understanding of the material covered in each section of the course, and the stakes are much lower, compared to university entrance exams. From Akiyama's (2004) and Taguchi's (2005) studies, it is possible to deduce that teachers tend to ignore the educational policy in not only teaching but also testing. While their studies reported the clear existence of negative washback in OC courses, what should be further investigated is teachers' perceptions of the gap between the goals of the educational policy and the social reality.

In short, it seems that CLT has not been adopted in classrooms. Therefore, the literature indicates that, under the strong influence of the high-stakes exams, a relationship between the national curriculum and classroom practice (teaching and testing) has not been established in this context. Therefore, given that high-stakes exams hold "the power over teachers" (Shohamy, 2004), since there has not been any crucial change in the university entrance examination system yet, we can postulate that negative washback still exists in OC classrooms both in teaching and testing. It is especially important to draw more attention to the relation between classroom-based tests and washback phenomena.

2.5 Summary

This chapter started with a brief overview of theoretical washback studies to provide the definitions of the phenomenon. Followed by empirical studies on both positive and negative washback, it reviewed key studies that established the concept of the important role of the teacher factor played in the mechanism of washback. Then this review of the literature demonstrated that a combination of social, cultural and educational constraints appear to hinder the implementation of the national curriculum in Japanese high schools. In particular, university entrance exams were largely considered as a key factor causing this long-lasting pedagogical predicament. Regarding discrepancy between the educational policy and the strong influence of university entrance exams, researchers reported that teachers leaned towards prioritizing the latter. In the Japanese context, however, little has been investigated about whether teacher-related variables, particularly schools' academic levels, have an impact on classroom practices. In addition, there is little research in Japan on the relationship between negative washback and classroom-based testing. Hence, the remainder of this thesis reports a study that builds on earlier work on negative washback in OC courses and will document EFL teachers' professionalism under the strong influence of university entrance exams.

Chapter 3: Methodology

3.1 Introduction

This chapter provides an overview of the methodology employed in the study. It begins with a restatement of the purpose of the study and the research questions. It then discusses the theoretical rational regarding the participants recruited, the research cites selected, the instruments and data collection techniques employed, as well as the analytic processes followed. It then makes an argument for the selection of a mixed methods approach for the study. Finally, the chapter describes in detail the four data collection instruments used, which were specifically designed to examine the central research topic. After presenting these one by one, it describes the research participants, the setting, and the data collection and analysis procedures.

3.2 Purpose of the Study

The main objective of this research was to examine EFL teachers' perceptions on classroom teaching and testing under the strong influence from a high-stakes test that does not match educational policy, since it lacks any interactive speaking component. In addition, this research aimed to identify influences on speaking-focused courses from teachers' perspectives. The term, *teaching and testing*, is used in this thesis to refer to classroom practice: what and how teachers teach and test/assess in the speaking-focused courses. Based on the premise that educational policy of English was formulated to facilitate and improve the students' language acquisition process, the Course of Study, classroom teaching, and classroom testing should share the same principle. The university entrance exams, similarly, should represent the tenets of the national curriculum content. However, although it is discussed theoretically, synchronization across the curriculum of teaching and testing has not been sufficiently explored in the literature, especially using empirical evidence.

In addition, the study addressed what teachers of the courses considered as influential factors on classroom practices. Although washback is claimed as a major influential factor of EFL education in Japan that hinders teachers from adapting CLT, especially at the high school level (Gorsuch, 2000, 2001; Hiramatsu, 2004; Kikuchi & Browne, 2009; Reesor, 2002; Taguchi, 2005), teachers' perceptions of its influence on in-class testing have not been fully studied yet. In particular, no study has investigated the issue across two different demographics: teachers in academically prestigious classes and those in less academic classes. Considering the point made in the literature that teachers' backgrounds are an important element of the teacher factor (Turner, 2006, 2009; Wang, 2010), this study deliberately looks at the students' academic levels, which can help understand teachers' working environment.

3.2.1 Research questions.

In this study, therefore, the following two research questions in relation to the above-mentioned issues are discussed:

1) How are teaching and testing in speaking-focused courses of English related to one another when the educational policy and the high-stakes exam are not in alignment?

2) What are the nature and the scope of washback effects from the university entrance exams on teacher perceptions in the speaking-focused courses?

In these questions, "Teaching" means what teachers do to conduct lessons of Oral Communication (OC) courses, and by "testing," this study refers to how teachers use low-stakes exams such as in-school and in-class exams to measure students' linguistic development of the target language. In addition, this study examined what teachers considered as influential factors on their teaching in OC courses and the reasons affecting such factors.

With a view to answering the research questions, the research instruments and the data collection and analysis procedure were specifically designed and the selection of participants and research sites was purposely undertaken. The rationale for using these instruments and the overview of the methodology are described in the following sections.

3.3 Methodology

3.3.1 Mixed Methods Sequential Explanatory Design

This study was conducted in the framework of mixed methods research (MMR) which combines two different data sources—quantitative and qualitative—to provide a more holistic picture of a research problem (Creswell & Plano Clark, 2011). The advantage of using MMR is its robustness "to explore the broad and complex set of research questions without the constraints associated with using a single method or technique" (Kingston, Sammons, Day, and Regan, 2011). Particularly, in this study, the mixed methods sequential explanatory design was employed. In the explanatory design, the primary emphasis is usually on the quantitative aspects that provide a general understanding of the research problem. The researcher initially collects and analyzes the quantitative data, and, in the second sequence, qualitative data are collected to help explain and build on statistical results by exploring participants' views in more depth (Creswell & Plano Clark, 2011; Teddlie & Tashakkori, 2009).

The rationale for this approach is due to the nature of the main issue under study, washback, which is largely recognized as a complex phenomenon (Alderson & Wall, 1993; Bachman & Palmer, 1996; Bailey, 1996; Cheng, 1997, 2000; Watanabe, 2004). Therefore, the distinctive characteristics of the MMR methodology are relevant to and compatible with the goals and characteristics of this study. Moreover, the explanatory design specifically suits the objectives of the study, which aims to provide a broad-spectrum perception of the current

situation of OC courses. A unique approach of this research design is that by integrating three different sources of qualitative data, the study attempts to explore richer and clearer insights into the findings emerging from the quantitative data (see Figure 2). Therefore, this methodology was used in order to investigate the aforementioned research questions. The research instruments developed for the study will be presented in detail in the following section.
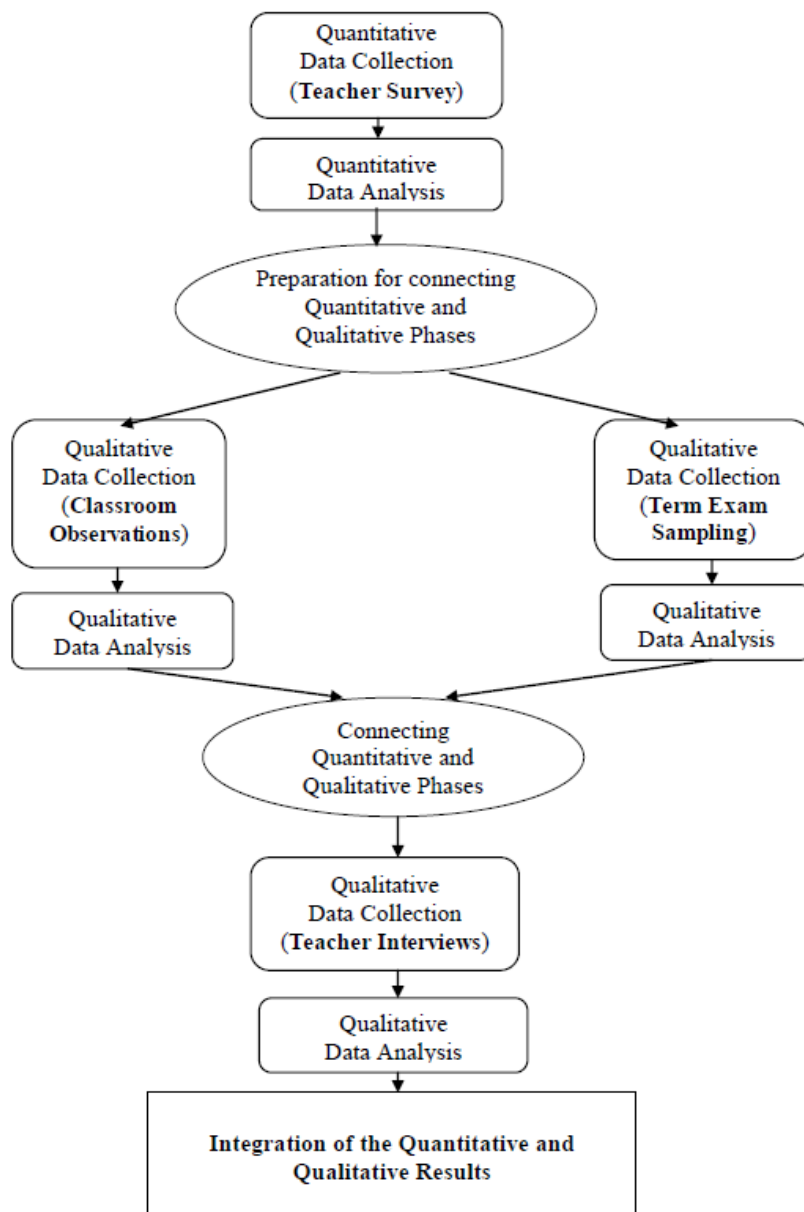
Figure 2 Visual diagram of the research design of the study

3.4 Access, School Profiles and Participants

3.4.1 Access.

  The study was carried out according to standard research ethics at McGill University. Data collection started in the end of April and was completed in the first week of June 2010 during the first academic term in Japan, which runs from April to July. The study was conducted in one prefecture where the researcher used to work as an EFL teacher. Prior to the data collection, in January 2010, I contacted the Board of Education to seek permission to conduct a research project in the region. After receiving official permission from the Board, I was able to access the schools and participants. All of the potential participants received a consent form and a letter that explained the objectives of the study at least two weeks before their participation, and they knew that they could withdraw from the study at any time without consequences.

3.4.2 School profiles: Academically-inclined vs. less academically-inclined classes.

  Along with the diversity of the interviewees' backgrounds, their working environments—their schools—can also distinguish their professional perspectives. In this study, the participants were recruited only from public high schools in one region. In general, however, each school has a different level of prestige in terms of the students' academic performance. What is more, even in the same school, each class is likely to be grouped depending on a student's needs and academic performance in Japan. Thus, it is of great importance to consider this particular point that makes classes display different characteristics. Hence, in this study, as a key characteristic of the teacher factor, the information about the academic level of each class was collected.

  The number of students planning on taking university entrance exams in a given class causes the teacher to change his or her teaching strategy/pedagogical approach. For example, the percentage of the students who write the annual standardized university entrance exam called the

Center Test (National Center Test for University Admissions) can be a major indicator of the school's academic profile. The Center Test is the achievement test for admission into all public universities as well as many private universities in Japan, and the annual event is tremendously high-stakes for all test takers. For that reason, in this study, the variable was analyzed statistically as a key factor and comparisons were made to see if there were any differences in teachers' perspectives between those in academically-inclined (AI) classrooms and those in less-academically-inclined (LAI) classrooms. In this study, an AI classroom is defined as a group in which more than 50% of the students will write the Center Test, and in LAI ones, less than 50% of the students are hoping to take the exam. In general, a school that sends many students to prestigious universities every year attracts academically strong students to apply to the high school. Thus, classrooms in such a high school tend to be all AI classrooms. Nonetheless, although it is not common, in a LAI school, there are classrooms for students who plan to study at the tertiary level. In this study, based on the collected information from the teachers, the classrooms was categorized into the two categories.

3.4.3 Participants.

With the aim of collecting data from teachers who were familiar with OC courses, the following criterion was set: the participants were EFL teachers working full-time in public high schools, who either a) were teaching an OC course at the time of data collection or b) had taught an OC course in the previous two academic years. On the premise that the study focused on the OC courses, information provided by participants who did not meet the criteria were eliminated from the data.

In the first phase of data collection, which consisted of a teacher survey, 87 teachers who met the aforementioned criterion responded.

In the second phase—classroom observations—five teachers agreed to participate. Although the participants were assigned by the Board of Education, I made a specific request to have participants working in different schools. The rationale behind this request was in order to observe if there would be any differences depending on the types of schools; in other words, in relation to RQ 2, I needed to observe lessons in different classroom environments. At that point, three teachers in AI classrooms and two from LAI ones were volunteered by the Board of Education.

Thirdly, the same five teachers from the second segment agreed to provide in-class term exam samples that could be compared to their classroom teaching practices. Due to the confidentiality of the in-school term exams, I obtained permission from the head teachers of the EFL department in each school to collect the samples.

For the last segment of the data collection, ten participants volunteered for the guided interviews: five were the same teachers who participated in the second and third phase so as to provide a more complete picture of their classroom practice, and five other teachers were recruited to obtain a broader view. However, an interview with one teacher who turned out not to meet the criterion was excluded from the data set. The nine teachers had diverse backgrounds, especially in terms of professional experience and personal experiences living in English speaking countries, which are reported in Chapter 4 (See Table 6 in Section 4.1.1.4).

3.5 Data Collection and Analysis Procedures

3.5.1 Instruments.

The data collected for this research come from four sources: the teacher questionnaire; the classroom observations; the term exams; and the teacher interview. By following the mixed methods sequential explanatory design, the first data source—the teacher questionnaire—was

used to produce the core qualitative data that helped illustrate the overall picture of the updated

reality in OC classrooms in this context. The qualitative data were used to explain and build on

the statistical results that emerged from the findings of the teacher questionnaire. The summary

of data collection is presented in Table 3, and the four instruments are attached in the

Appendices.

Table 3 *Summary of data collection*

Source of data	Participants	Instruments	Data analysis
1. Teacher survey	- 87 Japanese teachers of English in public high schools	questionnaire	quantitative
2. Classroom observations	- 5 teachers - 2 lessons from each teacher	classroom observation guide	qualitative
3. Term exams	- 5 term exams - 1 from the 5 teachers	term exam analysis guide	qualitative
4. Interviews	- 10 teachers (including 5 teachers from Segment 2 and 3, and 5 other teachers)	interview guide	qualitative

3.5.1.1 *Survey*.

First of all, as a primary source of quantitative data, a teacher survey (see Appendix A) was developed to identify how teachers organized lessons and term exams of OC courses. Some of the questions in the survey were adapted from Gorsuch's research on Japanese educational policy (2000, 2001) as they were highly relevant to this study (see Q9 in Appendix A). Targeting the same population—EFL teachers in public high schools—this study examined whether the influential factors identified in Gorsuch's study were applicable to a similar context ten years later, specifically in OC courses. The questionnaire was comprised of the following parts: 1) teacher's background information; 2) 4-point Likert scale questions regarding their OC course

organization; 3) 4-point Likert scale questions regarding assessment methods of OC courses; and

4) 4-point Likert scale questions regarding influential factors on the courses. The first part

included questions concerning their own background information (educational and professional)

and their working environments such as academic levels of their students.

3.5.1.2 *Classroom observations.*

Secondly, followed by preliminary analysis of the questionnaire responses, the classroom

observations were conducted to generate qualitative data to examine the specific activities and

teaching methods teachers were engaged in. To be more specific, the instrument for this segment

was used to mainly focus on the following aspects: 1) types of classroom activities; 2) focus of

lessons; 3) instructional use of English and Japanese; 4) teaching methods; 5) relevance to the

university entrance exams.

Given that the research questions set out to capture teacher classroom behaviour in detail,

an adapted version of the guidelines developed by Turner (2000) was employed. Based on

Turner's model, the classroom observation guidelines were modified to fit the specific target

context (see Appendix C). With this instrument, the data were documented manually at the site.

An integrated circuit (IC) recorder was also used to electronically save the data for further

analysis. Each observation lasted 50 minutes, and the data were collected from five teachers. The

teachers and their students were aware that I was observing the lessons for research purposes.

3.5.1.3 *Term exams.*

Thirdly, samples of in-school term exams of the courses, which were all paper-based,

were collected from teachers who participated in the second phase in order to juxtapose their

teaching practices with testing practices. For this study, the first mid-term exams were collected.

Table 4 represents an example of the assessment routine in Japanese public high schools. Term

exams are administered at each grade level at least once, and usually twice, per semester in

Japanese high schools (mid-term and term-end exams). The school-based exams are generally

designed as summative assessment tools and administered by teachers of the English department

in schools. In addition, in-class tests/assessment activities are generally managed by the teacher.

The in-class level tests are more context-bound, and managed by the teacher to meet the needs of

his/her class. Compared to university entrance exams, the stakes of internal tests are much lower.

To investigate the teaching-testing congruence in this context, based on my own

experience as an EFL teacher, I developed a term exam analysis guide to examine task types and

the focus of the term exam (see Appendix D). This instrument was designed to examine the task

types used in an exam and its score allocation to determine the weight given to each task by a

teacher-tester.

Table 4 *An example of a Japanese academic year and the basic assessment routine in public*
high schools

	Term 1	Term 2	Term 3
	April - July	September - December	January - March
In-class	routine classroom-based assessment		
In-school	**Midterm exam**	**Midterm exam**	**Final exam**
	Term-end exam	**Term-end exam**	
External	Achievement test	Achievement tests	Achievement tests
		(local and/or nation-wide)	The Center Test and
			university entrance exams
			(Grade 3)

3.5.1.4 *Interviews*.

Finally, face-to-face interviews were undertaken to help interpret and understand the emergent findings from the other sets of data. The design of the interview guide was informed by findings from the teacher survey, classroom observations, and term exams. In the guided individual interviews, open-ended questions were included in the interview protocol (see Appendix E) in order to elicit teachers' views on their classroom practices. The interviews were recorded with their permission.

3.5.2 Data collection procedures.

At the outset, the teacher questionnaire was sent to all Japanese teachers of EFL (approximately 300) working full-time in public high schools in the prefecture. For the classroom observations, in-class term exam analysis, and the face-to-face interviews, data were collected through on-site visits to high schools. Through the Board of Education, I contacted the schools and teachers who volunteered to participate in the study. Subsequent to classroom observations and term exam analysis, face-to-face interviews were undertaken to help interpret and understand the emergent findings from the previously collected data and to probe issues further. Therefore, interviews with teachers were conducted after the completion of the preliminary analysis of the survey, classroom observations, and term exams. In the guided individual interviews, open-ended questions were employed to record teachers' opinions. Each interview lasted about 30 minutes and was conducted at the high schools where the participants worked. With the consent of the interviewees, all interviews were digitally audio-recorded. The teachers were given the option of doing the interview either in English or in Japanese. All of the participants chose to be interviewed in Japanese.

3.5.3 Data analysis.

As described above, the data were collected and analysed within the framework of a mixed methods sequential explanatory design. First of all, data from the teacher survey were analyzed as the primary source of quantitative data. A statistical analysis software application, PASW Statistics 18, was used to analyze the quantitative data and to interpret the effect of the school's academic level variable.

Secondly, the data from classroom observations were coded and labelled to examine the focused aspects of teaching practices. The audio-recorded data were transcribed, using a software application designed to play back audio recordings for transcription. The software, Express Scribe Version 5.10, was used also for the interview transcription. Thirdly, using the analysis guide I developed for this purpose, samples of in-class term exams were used to obtain quantitative and qualitative data to describe what types of tasks teachers used to measure students' English skills in the speaking-focused courses. Finally, the interviews with teachers were analyzed as a primary source of qualitative data to provide richer, deeper, and clearer insights into the quantitative findings. The recorded data were transcribed, coded, translated, and paired with relevant survey questions, using the aforementioned software. Because the interviews were conducted in Japanese, the transcribed text was translated into English for reporting after the data analysis was completed.

3.6 Summary

The methodology used in this study exhibits characteristics of both quantitative and qualitative approaches combined, that is, a mixed methods research design. The survey participants were 87 EFL teachers working at public high schools in one specific prefecture in Japan. For the classroom observations and term exam sampling, five teachers volunteered. They

also participated in the interview segment along with four other teachers. Thus, in total nine teachers provided their opinions in the guided-question interviews.

I designed and organized this sequential use of the four instruments specifically for this study, and the questionnaire and the classroom observation guide were adapted from Gorsuch's (2000) survey and Turner's (2000) classroom observation guide for their relevance to the study. Likewise, the analysis guide of term exams was developed for the study. The interviews were conducted in an open-question manner so that the teachers could discuss the issues candidly. Each instrument was vital to illustrate the whole picture of the curriculum-testing congruence in OC courses.

Having presented the research methods and procedures, the following chapters will turn to discussions of major research findings emerging from the data.

Chapter 4: Presentation of the Findings

4.1 Introduction

The results of the study are presented in this chapter guided by the following research questions:

RQ1: How are teaching and testing in speaking-focused courses of English related to one another when the educational policy and the high-stakes exam are not in alignment?

RQ2: What are the nature and the scope of washback effects from the university entrance exams on teacher perceptions in the speaking-focused courses?

To address the research questions, the data were analyzed and organized around three main themes. Firstly, teaching practices: what teachers do during OC classes, which will help illustrate the organization of the courses. Secondly, testing practices: what and how teachers assess students' abilities, which will provide concrete information to discuss whether classroom teaching and testing/assessment are in alignment. Finally, influential factors: to what extent university entrance exams are considered as barriers to the implementation of the educational policy in OC courses. Moreover, what kind of other constraints teachers consider as influential factors. The results will be presented in the form of the comparisons of AI classes and LAI classes.

To address the first theme, data generated through the teacher survey, the classroom observations, and the teacher interviews were used. Next, for the second theme, data from the teacher survey, the term exam samples, and the teacher interviews were analyzed. Then, to present the last theme, data from the teacher survey and the interviews were used.

4.1.1 Regarding the instruments and profile of the participants.

4.1.1.1 *Questionnaire*.

Eighty seven teachers responded to the survey (27.4%). The final part of the questionnaire dealt with their background information and teaching environment. Figure 3 presents the general characteristics of the survey participants. All teachers reported that their first language was Japanese. This reflects the current Japanese policy that does not permit non-Japanese people to work as full-time employees in public schools. Five participants (6%) reported that they had an experience of living abroad for more than one year. In terms of their university degree level, 72% of them held a BA degree and 11% had a master's. The respondents' university degrees had predominantly focused on the following three majors: English literature (32%), English linguistics (24%), and Education (20%). Merely three teachers (3%) reported having received teacher training in TESOL, and thirteen teachers (15%) were from other majors such as anthropology, economics, business, psychology, and Russian and German literature. One important fact that emerged concerning their training experiences is that very few teachers had received training on CLT (pre-service 10% and in-service 8%) although the instructional approach had been promoted by the government for decades.
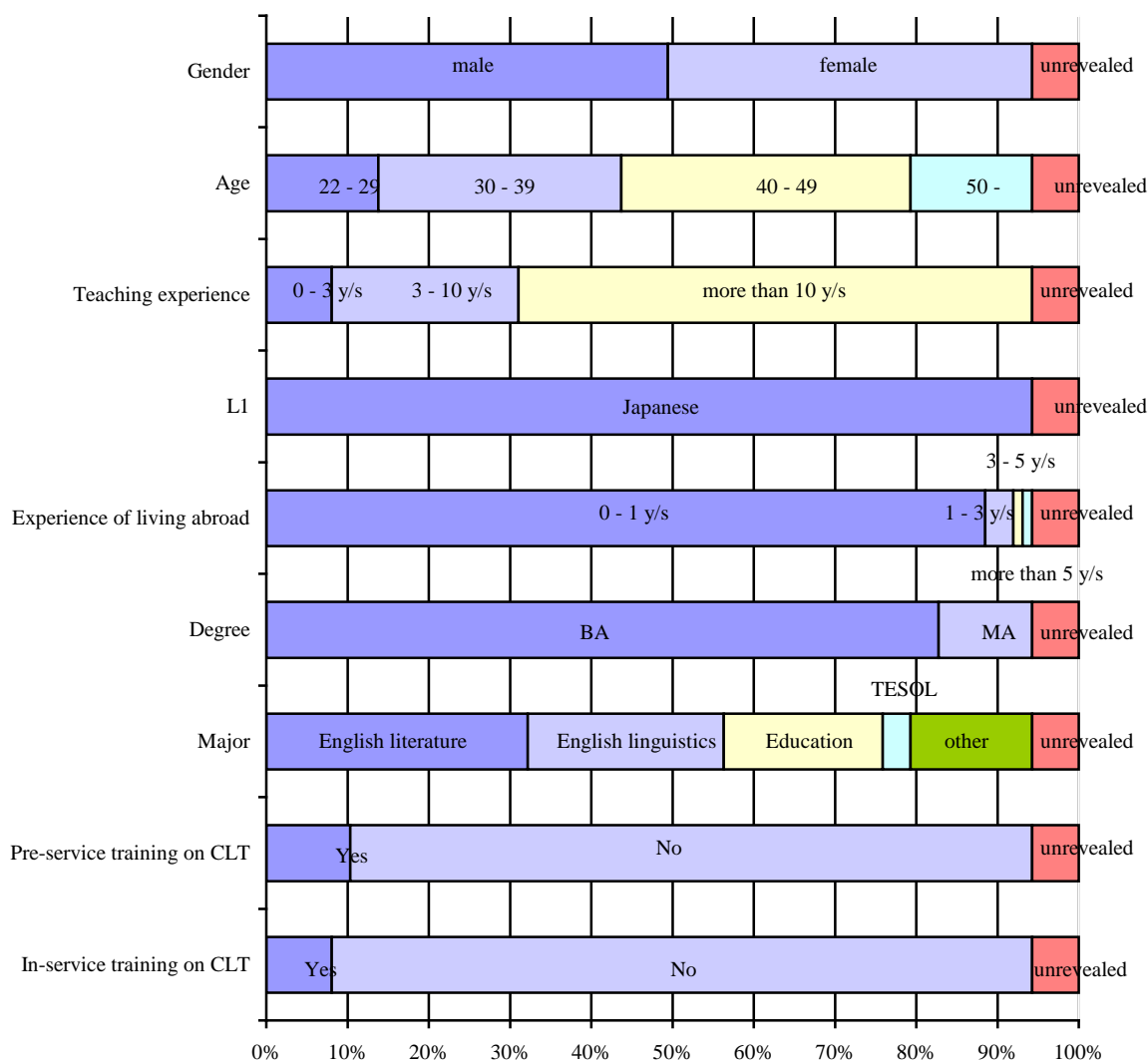
Figure 3 Background information on the survey participants

4.1.1.2 *Classroom observations*.

The initial purpose of classroom observations was to juxtapose the qualitative data of the

survey with the reality of classrooms; however, one teacher, during her interview, disclosed that

she had manipulated her lesson due to the presence of the observer. According to her, she

allocated more time for speaking activities to present a more 'Oral Communication-like' lesson;

therefore, the data from her lesson were eliminated from this study. This influence, what Labov

calls the "observer's paradox" (Labov, 1972, p. 209), may be applicable to the other four

teachers' lessons as well, considering the fact that all of the teachers were assigned by the Board

of Education and were aware of the research objectives. Hence, in this study, the data obtained

from classroom observations were not used to draw a general view. However, some noteworthy

facts observed in common among the four teachers should be presented.

       Table 5 represents the description of the observed lessons. All four classes were taught

by different teachers working in four different schools, and the classes displayed different

characteristics such as class-size, grades, and academic prestige. To protect the identity of the

teachers who volunteered, pseudonyms were given to all participants. For instance, Class A,

taught by Mr. Nishida, is defined as AI class, as potentially all students were going to take the

Center Test at the end of their high school life. Class C then falls into a group of less-

academically-inclined classes, most of whose students do not take the high-stakes exam.

However, despite these differences, the four classrooms demonstrated some similarity in their

activities.

Table 5 *Description of the observed classrooms*

Class	Teacher	Courses	Grade	Class -size	Percentage of students who advance to higher education	Percentage of students who take the Center Test	Academic prestige
A	Mr. Nishida	OC1	1	40	100%	100%	academic
B	Ms. Toyama	OC1	1	40	91~100 %	71~90%	academic
C	Ms. Kurano	OC 2	2	17	0~10%	0 %	less-academic
D	Ms. Hori	OC1	1	30	100%	100%	academic

4.1.1.3 *Term exams*.

Five samples of term exams were collected from the teachers who participated in the classroom observations. As mentioned above, data provided by one teacher were eliminated; thus, four term exams were analyzed.

To collect data, teachers were asked to describe how they designed their term exams with an actual sample of their term exams. Ms. Hori, who taught Class D, agreed to provide a real term exam that was designed by a colleague for the school. However, unfortunately, the other three teachers had not finished preparing the test yet, because the data collection was done a week before the mid-term examination. Therefore, instead of having a printed term exam in front of us, they orally explained the design of the term exams in detail. With the term exam analysis guide, the teachers described how they allotted scores to what kind of tasks.

4.1.1.4 *Interviews*.

Nine teachers who participated in the interview sessions reported their candid opinions on the OC courses. Table 6 presents the teachers' background and working environment information. Again, pseudonyms were given to the interviewees, and alphabetical codes were allocated for the name of the class and the institution in order to protect the individuals' identity. All interviewees were EFL teachers working in five different public high schools in four different cities. Of the nine participants, four teachers were less experienced, with teaching experience ranging from five to seven years, and the other five teachers were highly experienced professionals working as EFL teachers in the public school system more than seventeen years. Among them, two teachers had an MA degree in TESOL, and the other seven teachers were from either English literature, English linguistics, or English education. One teacher who held an MA degree from a US American university had had the experience of taking an assessment course;

however, the other eight teachers reported not having had any experiences of training in

language assessment.

Table 6 *Description of the interviewees*

Name	School	Gender	Years Teach.	Study degree(s)	Major	Living Abroad.	OC	% of Ss taking CT in the class
Nishida	A	M	6	BA/MA	Economics / TESOL	1 yr.	1	over 90%
Toyama	B	F	7	BA	English Linguistics	1 yr.	1	71-90%
Kurano	C	F	6	BA/MA	English Education/ TESOL	2 yr.	2	0%
Hori	D	F	20	BA	English Literature	1 mo.	1	over 90%
Satou	E	F	5	BA	English Linguistics	0	1	31-50%
Yamada	B	M	20	BA	English Literature	1mo.	1	71-90%
Abe	C	M	18	BA	Linguistics	0	1	10%
Inoue	D	F	24	BA	English Literature	0	1	over 90%
Mizuno	E	F	17	BA	English Education	6 mo.	1	31-50%

4.2 What Teachers Teach in OC Courses

The findings revealed that the Course of Study had some influence on the teaching

practices. The collected data were organized by following the two areas. One is about classroom

organization and activities, and the other one is concerning teachers' use of English to conduct

lessons.

4.2.1 Classroom organization and activities.

4.2.1.1 *Questionnaires*.

Regarding their teaching practices in OC courses, first of all, the data revealed that the

class size was still large, in excess of 30 on the average (see Table 7), which reflects the reported

reality of EFL classrooms in Japan (Gorsuch, 2000; LoCastro, 1989). The large class size has often been proposed as a key obstacle to the improvement of the EFL education in Japan; yet, this recent information shows that the status quo still prevails.

Concerning what teachers do in OC courses, the Course of Study (MEXT, 2003) repeatedly emphasizes that listening and speaking activities should be the key components of lessons. Amongst the listed activities in Table 8, the most highly rated activity was listening exercises. In light of the congruence of course objectives and teaching, the results of three activities, chorus reading, listening exercises, and pronunciation/accent reduction activities, may be a sign of teachers following the course objectives. Especially, as the course objectives state, listening is one skill that teachers should focus on in OC courses. Moreover, the other main skill to focus on—speaking—also seemed to be taught frequently in classrooms. In fact, 54% of the teachers regularly engaged students in pair or group work activities, while individual presentations in English ranked as the least popular activity. With reference to opportunities for pushed output, it is noteworthy that 52% — more than half— of the teachers stated that their students had never had the opportunity to present individually. Additionally, the teachers reported that grammar-focused lectures were still common; 46% of them answered that they often or always allocated class time to the lectures.

Table 7 *Number of students in class*

Mean	*Mode*	*Median*	*SD*	*Min*	*Max*	*Range*
31.52	40	38	10.68	7	40	33

Table 8 *Activities used in the classroom*

Description	1. never	2. sometimes	3. often	4. always	*M*	*SD*
Chorus reading	5%	15%	33%	47%	3.23	.87
Grammar/Vocabulary quizzes	16%	41%	37%	6%	2.32	.81
Grammar focused lectures	18%	36%	35%	11%	2.39	.92
Linguistic/Cultural lectures	9%	49%	41%	1%	2.34	.92
Pronunciation/accent reduction	2%	28%	39%	31%	2.99	.83
Listening exercises	3%	18%	29%	50%	3.24	.88
Pair/Group activities in English	14%	32%	30%	24%	2.64	1.00
Individual presentations in English	52%	36%	12%	0%	1.68	.70

4.2.1.2 *Classroom observations*.

In the observed classrooms, the major activities to which the teachers allotted the most time were listening exercises and text translation from English to Japanese. In addition, although all teachers used speaking activities in their classrooms, the purposes of the tasks were not clear and they were organized in a highly teacher-centered manner.

First of all, consistent with the results of the survey, all four teachers allocated considerable amounts of class time to listening comprehension exercises. The teachers operated audio-devices, such as CD/MP3 players, to administer listening activities, utilizing recorded materials created by publishing companies specifically for listening exercises. During those activities, seldom did interactions of either teacher-student or student-student take place, since

students were working individually to answer multiple-choice questions. The following dialogue

is an excerpt from Class D.

> *Now please check your answer(s) by yourself. No.1*… えーと、Aにした人いますか？A
>
> にしたひと、はい、*You are right.*
>
> [Translation: Now please check your answers by yourself. No.1…well, are there anyone
>
> who chose A? Those who chose A, yes, you are correct.] (Ms. Hori, Classroom
>
> observation: 26/05/2010)

Even after the activities, any meaning-focused interactions in the target language, which might

elicit students' pushed output, were not observed.

Another observable characteristic was popularity of pattern practice as a speaking activity.

Using a dialogue in the textbook, paired or grouped students practiced vocalizing the given

sentences a few times. In other words, during the 'speaking' activity, the students were not

engaged in meaningful language use but in merely mechanical practice. Exceptionally, the

teacher in Class B used an information gap activity as a warm-up task for a few minutes, in

which a student had to describe a given object in English without using its actual name. However,

the activity lasted for a few minutes without any feedback provided by the teacher.

Lastly, language-focused activities were also a common characteristic among the

classrooms. Language-focused learning is, as Nation and Newton (2009) define it, "learning

through deliberate attention to language items and language features" (Nation & Newton, 2009).

Decontexualized language tasks such as vocabulary, grammar, and translation seemed very

common. Translation, in particular, was the central component of the lessons in which teachers

paid much heed to accuracy of the Japanese translation, and the students were to translate a text from English to Japanese, but not from Japanese to English. In Class A and D, the teachers provided detailed metalinguistic information in Japanese to enhance accuracy of translation. As an example, the following excerpt from Class A shows the teachers' explicit provision of metalinguistic information:

時間が懸かるっていう表現は他にも、*take* っていうのを習ってるはず。この前のリスニングパイロットの１でやったね。その場合の主語はなんだったか。*it* だったよね。*It spend* って言えないんだよ。*It takes*、人、 時間だったね。

[Translation: You have already learnt another expression to say 'spend time'. We studied 'take' in the previous lesson on the textbook, Listening Pilot, in Part 1. What should you use for the subject coupled with the verb 'take'? It's 'it'. You cannot say 'It spend(s)'. 'It takes', and 'someone', then 'time' follows.] (Mr. Nishida, Classroom Observation: 05/07/2010)

During the grammar-focused lectures, the students were attentively listening to the teacher and taking notes. Likewise in Class D, the teacher devoted half of the class time to focus on grammar drills in which, again, any meaning-focused interaction in the target language was not observed. Furthermore, for the grammar-focused lecture, they used a sub-textbook that was, namely, a grammar exercise book, besides the main OC textbook authorized by MEXT.

4.2.1.3 *Interviews*.

According to the interviewees, they considered oral/aural activities as an important component of OC courses. They described how they organized every lesson in an attempt to

integrate listening and speaking activities effectively. On the other hand, they also mentioned

that it was highly unfeasible to provide ample opportunities to let the students practice oral/aural

skills in the limited class time available. The following excerpt, from an interview with Mr.

Nishida, provides an example of the situation where he organized his OC lessons with

frustration:

リスニングって力をつけるには相当な量を聞かないといけないと思うんですよ。相当な
量。それを、えーと、週に２コマくらいのオーラルの授業で１０分程度、スクリプトで
言ったら、ほんの１分にも満たないようなスクリプトを聞いたところでどうなるのって。
（中略）なんぼこの学校の授業の中でリスニングの量をあげたところで、そんな聞ける
ようにならんと思うんですよ。（中略）それをするよりも読み書きをやっていったほう
がいい。

[Translation: To develop strong listening skills, I believe you need to devote considerable

time to the activities. An astronomical amount of time. Well, then, what on earth can we

expect from students by providing 10 minutes listening activities, which could be just less

than one minute input, within mere two lessons a week? …No matter how much we

increase the amount of time for listening in lessons, I do not think [students] will be able

to develop the skills…It is better to teach reading and writing than doing such a thing.]

(Mr. Nishida, Interview: 05/17/10)

As well as Mr. Nishida, Ms. Toyama, who only used an interactive speaking task that pushed

students' output, revealed her real intention of including the task in her lessons. She told me that

her focus in the OC lessons was on listening, not speaking, due to the importance of university

entrance exams. Exceptionally, her lesson was conducted mostly in English, and she stated that it was to increase the amount of input to develop students' listening skills. However, regarding the purpose of the speaking activities, she did not aim at the improvement of students' speaking skills but at "a diversion, a review, and a bit of practice." She said:

> 本当のスピーキング力をつける、付けたいがためのスピーキング活動ではないですね。
>
> [Translation: It's not activities in order to develop speaking skills, their real speaking
>
> skills.] (Ms. Toyama, Interview: 05/19/10)

Likewise, other teachers expressed their recurring doubts about whether it was possible to help improve the fluency of students' speaking skills in OC courses.

The teachers also voiced that they still valued language-focused activities such as traditional grammar and translation drills, and that those activities should be included in lessons of OC courses in order to enhance students' accurate understanding of English. In addition to that reason, teachers emphasized that the language-focused activities were of great necessity for students who would write the university entrance examination. In Mr. Nishida's opinion, teachers needed to primarily develop students' literacy in English, even in OC courses, because the chief incentive for their learning experience was to pass university entrance exams. Therefore, he utilized the grammar-translation method in his classroom, albeit questioning himself occasionally during his lessons, as he reported in the interview. Similarly, other teachers confessed that with the pressure of preparing their students for the exams, they often had to abandon time-consuming interactive activities that would seldom affect their students' future as test-takers.

4.2.2 The use of instructional languages.

4.2.2.1 *Questionnaires***.**

In terms of the use of English in the classroom, as Table 9 indicates, with a mean of 2.16, the main language used as the instructional medium remained Japanese. It is of great importance that only 3% of the 87 teachers in the survey reported using English to conduct lessons in "Oral Communication" courses; moreover, 10% of the teachers reported they did not use English in the classroom at all.

Table 9 *Use of English to conduct lessons*

1. never	2. less than half	3. more than half	4. mainly in English	M	SD
10%	67%	20%	3%	2.16	0.64

4.2.2.2 *Classroom observations***.**

As mentioned, the participants, some of whom had had the experience of living abroad for an extended period, seemed to be quite comfortable with their English speaking abilities. All of them, in fact, interspersed English throughout the lessons. Nevertheless, by and large, the main medium of instruction was Japanese in the observed lessons, and English functioned only as a classroom management tool to organize classroom work in a one-way manner that did not require responses from students. In addition, the teachers' use of the two languages demonstrated a peculiar tendency; that is, they switched from English to Japanese frequently to translate what they said in English. The following excerpts exemplify the nature of the use of instructional languages in the classrooms:

You have to take turns. 交代してくださいね。で、交代して２回どおり終わったら、座りなさい。

[Translation: Then, you can take a seat after practicing two times by taking turns.] (Ms. Hori, Classroom observation: 05/26/2010)

Later, you will do No.2. So... you have to take notes. You should take notes. メモを取ってください。空いているところ、どこでもいいから。

[Translation: Please take notes… wherever available.] (Ms. Toyama, Classroom observation: 05/11/2010)

Generally, teachers' utterances in English were coupled with Japanese translation, especially when these were long. Consequently, rarely did the students have opportunities for spontaneous language production in English, as they responded in the L1.

4.2.2.3 *Interviews.*

Even for teachers who reported that their lessons were largely conducted in the target language, they showed hesitation about the "English-only movement," considering what the consequence would be for students who would take the university entrance exam. For instance, Ms. Hori, who was in charge of Class C, asserted that CLT appeared unfeasible to prepare her students for university entrance exams, due to their students not being able to follow all instructions she gave in English. In her opinion, an advantage of learning from a Japanese EFL teacher was that students could obtain a deeper understanding of grammar through L1, which was crucial for her students. Therefore, she chose Japanese as an instructional language even in OC courses.

In the case of a less academically-inclined context, however, Ms. Kurano also revealed her uncertainty about the use of English as a medium of instruction in her classrooms. She stated that, for the sake of classroom management, she interspersed Japanese throughout her lessons so that the students would not be confused but efficiently engaged in any activity. In her case, none of her students were going to write the Center Test, so for smoother classroom management, she chose using Japanese to make her instructions clear. In fact, nearly all teachers, including Ms. Kurano, emphasized that the goal of the courses was "to foster a positive attitude toward communication" (MEXT, 2003, section II, para.1), which is the key element of the course objectives. This phrase seemed to be engraved in their minds and they considered that conducting lessons exclusively in English would harm the students' positive attitude toward communication owing to their inability to understand the instructions.

4.3 What and How Teachers Test in the OC courses

The findings revealed that the Course of Study had little influence in the testing practice in this context. In particular, there was a significant dearth of speaking tests in classrooms. In this section, issues of the incongruence with teaching and lack of speaking tests are discussed.

4.3.1 Incongruence with teaching.

4.3.1.1 *Questionnaire*.

Findings indicate, as shown in Table 10, that listening comprehension tasks ($M = 3.26$) largely appeared in the exams, which aligns with the results of classroom teaching. In term exams, the language-focused task evidently showed its enduring popularity. More than half of the participants inserted translation and grammatical tasks into the exams of OC courses. On the other hand, two task types, speaking ($M = 1.25$) and linguistic/cultural knowledge ($M = 1.68$),

were used extremely infrequently. Clearly, this indicates that the term-exams were not designed

in congruence with the course objectives.

Table 10 *Task types in school-based term exams*

Description	1. never	2. sometimes	3. often	4. always	*M*	*SD*
Reading comprehension	8%	46%	26%	20%	2.57	0.90
Spelling/Vocabulary	1%	16%	40%	43%	3.24	0.76
Grammar	10%	42%	23%	25%	2.63	0.98
Translation	5%	33%	36%	26%	2.84	0.87
Linguistic/Cultural knowledge	43%	48%	8%	1%	1.68	0.67
Pronunciation/Stress	5%	20%	36%	39%	3.09	0.88
Listening comprehension	7%	17%	18%	58%	3.26	0.98
Speaking	85%	9%	1%	5%	1.25	0.70

4.3.1.2 *Sample term exams*.

The four term exams collected from the participants revealed that they were in agreement

with the results of the survey. As noted earlier, listening was popularly taught in classrooms of

OC courses; however, in sampled term exams, the score allocated to listening tasks were not

high despite the fact that listening should be treated as a focus skill of the course. On the contrary,

a greater weight was granted to language-focused tasks in the paper-based exams. For example,

in the term exam of Class A, the teacher allotted 60% of the total score to tasks that tested

translation or grammatical knowledge. Writing skills were also measured in some tests, although

it was not observed in the classroom observations. What is more, all of the term exams included

reading comprehension tasks. These findings are in line with the teacher survey, and, most

importantly, the samples confirmed that speaking and linguistic/cultural knowledge were completely excluded from the tests, which was not in conformity with the teaching practices or the course aims.

Table 11 *Sample term exams*

Class	Focus of the question									scores allocated for sub-textbooks
	pronunciation/accent	spelling/ vocabulary	reading comprehension	listening comprehension	speaking	writing	translation	grammatical knowledge	sociolinguistic knowledge	
A	0	20	10	10	0	0	30	30	0	40/100
B	5	15	10	15	0	15	10	30	0	not disclosed, but sub-texts included other speaking tests existed for 50% of the final grade
C	0	15	35	0	0	40	0	0	0	
D	5	13	8	14	0	5	6	49	0	71/100
Mean	2.5	15.8	15.8	10.0	0	15.0	11.5	27.3	0	

4.3.1.3 *Interviews.*

These paradoxical gaps between teaching and testing were explained through the interviews. In order to design the term exam, teachers used not only the official textbooks but

also sub-materials such as grammar drill or vocabulary building workbooks. As a significant fact, except for the exam of Class C, the three teachers disclosed that questions in their exams were comprised of different materials, including the official textbooks. Therefore, the mismatch between teaching and testing was created in this context. Ms. Inoue described the complex design procedure of term exams in her school: "The students have four books for the courses. I think it is a bit overwhelming. They have to review all materials, not only the assigned textbook." Likewise, the other participants admitted to pulling questions from sub-textbooks to integrate into term exams in OC courses so that students would focus more on those language-focused tasks. "Since the very first day of their high school life, preparation for university entrance exams has started," said Ms. Inoue.

In the survey, the most highly rated classroom activity was listening exercises, which was somehow reflected in the term exams. However, interviewees imparted that the emphasis on listening skills was not generated by the educational policy but by the influence of university entrance exams. Mr. Nishida affirmed that the introduction of a listening component into the Center Test since 2006 had strong influence on his teaching practice, alluding to the washback effects. In this regard, Ms. Toyama also revealed what she considered to be the main priority in designing her term exams as follows:

いずれ模試、いずれ入試で役に立つ問題化っていうのを考えて（考査を）作るようにしているので、ＯＣの目標って考えたことが無いです。リスニングの問題は入れましょうってくらいですね。でも結局それは出来ないと入試に対応できないっていうのが頭の中にあるから入れてるだけで、ＯＣの目標に当てはめて入れているわけではないです、正直なところ。

[Translation: I always design my term exams, considering usefulness to take mock exams or university entrance exams in near future. I have never considered the course objectives [to create a term exam]. At least I do is including listening tasks, but because I am aware that listening tasks also appear in university entrance exams, and that's why I include it after all. To be honest, the purpose of the inclusion is not to meet the course objectives.] (Ms. Toyama, Interview: 05/19/2010)

As evidenced in her utterance, the influence of university entrance exams was mentioned by interviewees as a key relevant factor to design term exams.

In addition, concerning the curriculum-test congruence, when I asked the following question "do you think your term exams match the course objectives?," all interviewees answered immediately, "no", with a confused smile. For example, Mr. Nishida responded to the question as follows:

一致はしていないでしょうね、結局書かせてますから。定期考査は筆記試験ですから。えーと、挨拶とか、そのシチュエイションベースドの表現なんかに関しては、目標は達成させていると思います。そういう表現を覚えさせて、書かせたり選択させたりするような問題は必ず入れるようにしています。ただ、聞く、話すに関しては、まぁ聞くは１０点分入れてますから聞く力ってのは試していると思います。話すは…。

[Translation: It does not match. After all, it is written-based. Term exams are paper-based. Well, expressions like greetings or other situation specific ones, I think those accomplish the purpose [of the course objectives]. I definitely include those questions that students need to memorize or choose key expressions. But, regarding listening and speaking, well,

listening, because I include the task for 10 points, I believe I do assess their listening

skills by that. But speaking…] (Mr. Nishida, Interview: 17/05/2010)

He also told me that he would like to prioritize reading and writing, suggesting that speaking had

a low priority among the four skills for the sake of his students' needs.

A finding emerging from the interviews is that the teachers did not create term exams in

parallel with their classroom teaching. They also confirmed that term exams were used to

prepare students for the high-stakes exams. Moreover, the interviewees did not think testing and

teaching should be consistent with the course objectives.

4.3.2 Absence of speaking tests.

4.3.2.1 *Questionnaire*.

Without setting a performance test, how teachers are measuring the students' speaking

ability remains enigmatic. Table 12 presents the frequency of speaking tests in OC courses aside

from the term exams. Fifty-five percent of the teachers acknowledged that they did not allocate

any opportunities to test the students' speaking ability. Therefore, although some efforts to create

interactive activities were observed, especially pair/group work activities ($M = 2.64$), measuring

the development of students' speaking skills apparently was not at the centre of the teachers'

interests.

On top of that, even though speaking tests were administered in classrooms, seldom were

the scores reflected in the actual grades. On average, these tests only had a 7.14% weighting on

the final course grades (see Table 13). Again, lacking a component of speaking, these results of

testing practices do not match the teaching practices described above nor the course objectives of

OC courses.

Table 12 *Frequency of speaking tests per year outside term exams*

1. never	2. 1 or 2 times	3. 3 or 4 times	4. more than 5	*M*	*SD*
53%	29%	10%	8%	1.74	0.95

Table 13 *Percentage of speaking tests in the final grades*

Mean	*Mode*	*Median*	*SD*	*Min*	*Max*	*Range*
7.14	0	0	13.0	0	70	70

4.3.2.2 *Interviews*.

Except for the paper-based term exams, how and to what extent teachers measure students' speaking skills were also examined through the interviews. From what the teachers described, although speaking activities were administered in their classrooms, they did not test whether the activities helped students develop their oral proficiency or not. This result also matches the survey result, in which most teachers did not provide any opportunities for their students to be tested.

Most of the interviewees answered that they did not assess speaking at all in the courses. Some teachers stated that they could measure speaking skills from students' written production in English. As Mr. Nishida said above, tasks that involved reading or writing written dialogues could replace speaking tasks by testing their knowledge of how to respond in a correct manner in the target language.

As a noteworthy fact, moreover, some conceded that lack of confidence in assessing speaking abilities of the target language made them shy away from it. For them, the performance assessments were highly elusive in terms of accountability. "It could be very subjective and lacks validity," stated Mr. Yamada, displaying reluctance to actively integrate speaking tests into his

classroom practices. Similarly, some other teachers revealed their hesitation to judge speaking skills, and mentioned they were likely to avoid such an opportunity. No participants in this study had received any teacher training in the area of language testing, which might explain teachers' perspectives on the classroom test construction.

On the other hand, it should be noted that Ms. Kurano, the teacher of Class C, exceptionally told me that she gave speaking tests several times that were worth 50% of the final grade. Her assessment methods seemed to be tailored to the students' level, but in general, she usually provided model examples before the test so that her students would be able to practice or memorize them in advance. Ms. Toyama, the teacher of Class B, also mentioned that her students would have speaking tests at least two times. The two teachers referred to the Course of Study to express their testing philosophy. For instance, Ms. Kurano described her assessment method of an individual presentation task, explaining that the ultimate purpose of the task was to "develop the students' positive attitude towards English." Thus, her focal points of assessment were attitude, voice-loudness, eye contact, and quality of the content; that is, her grading rubric merely focused on voice and non-verbal communication. Another teacher, Ms. Toyama, also asserted that she valued students' kinetic messages more than verbal ones, believing that strict attitudes toward students' mistakes would harm their intellectual development. The following excerpt came from her interview in which she described her assessment strategy of a speaking test:

（評価のポイントは）きちんと準備をして当日自信を持って発表できるところまでやっているかどうか、です。例えば、スクリプトを原稿を手に持って棒読み…読んでいると、それは準備不足で三角ぐらいですね。で、これ紙を時々見るのも、まぁ、基本的にはス

ピーチですから、周りを見ながら顔を上げて、アイコンタクトをとりながらきちんと大

きな声で…だと丸。まぁ大きな声で堂々とはきはきと準備をちゃんとして、練習をちゃ

んとしてってあたりで。

[Translation: My assessment criterion is whether students are well-prepared for their

presentations on the day. For example, if a student holds a script in his hands to read it

out, then I give a triangle (so-so) because of the lack of preparation. Then, well, if he has

a sneak look at the script a few times, ah, you know. It is basically a speech test, he needs

to look around with his face up, making eye-contact, and in a decently loud voice. Then it

is good. So, [I assess students based on] like a loud voice, confident attitude, enough

preparation, good practice, those kinds of stuff.] (Ms. Toyama, Interview, 05/19/2010)

Apart from the fact that she equated a read-aloud test with a speaking test, what the test

measured was students' attitudes but not the quality of their oral production. This trend,

evaluating the student's "attitude toward communication" but not "accuracy" or "fluency," was

confidently emphasized by teachers. Although when it comes to paper-pencil tests, accuracy was

meticulously examined, the teachers contended that measuring students' accurate use of spoken

English would simply discourage them.

In short, most of the interviewees never assessed their students' speaking abilities in the

OC courses due to the uncertainty about their own assessment skills. Even in the few exceptional

cases, those who tested students' oral proficiency seemed to focus on students' attitude but not

on their oral production.

4.4 Influences on OC Courses

To provide a comprehensive outlook on this issue, this study explored possible barriers that were

preventing teachers from implementation of the national guidelines. As presented in Table 11,

teachers seemed to consider "students' insufficient English abilities" ($M = 3.26$), "term exams

and regional mock exams" ($M = 3.14$), and "class size" ($M = 3.06$) as the top three major

constraints on their decision making behaviour in the OC courses. "University entrance exams"

($M = 3.01$) ranked fourth, but collected the highest rate of "strongly agreed" (41%). Teachers

reported that training experiences, both pre- and in-service ones, had little influence on their

classroom practices.

Table 14 *Influential factors perceived by teachers*

Description	1. strongly disagree	2. disagree	3. agree	4. strongly agree	*M*	*SD*
Curriculum (the Course of Study)	5%	46%	41%	8%	2.53	.71
University entrance examinations	6%	29%	24%	41%	3.01	.97
Term exams/Regional mock exams (*moshi*)	0%	18%	50%	32%	3.14	.70
Pre-teacher training experiences	22%	52%	25%	1%	2.06	.72
In-service teacher training experiences	11%	54%	34%	1%	2.25	.67
Students' insufficient speaking skills in English	1%	11%	48%	40%	3.26	.71
Teachers' insufficient speaking skills in English	8%	20%	52%	20%	2.84	.83
Number of students in class	5%	21%	37%	37%	3.06	.88
Textbooks	5%	38%	34%	23%	2.76	.86

In addition, using the data from the questionnaire, an Independent Samples of t-test was then performed to test the equality of the means of the two groups. Statistically significant results were obtained in the means differences between teachers in LAI classes ($n = 47$) and ones in AI classes ($n = 40$). Among the nine variables, two variables—"university entrance exams" and "students' insufficient English abilities"—are the only tests that differed significantly between these two groups (Table 15). Regarding the factor of university entrance exams, the result of t-test shows that teachers in AI rated it higher ($M = 3.38$, $SD =.740$) than ones in LAI classes ($M = 2.70$, $SD = 1.041$) conditions; $t(85) = 3.417$, $p = .001$. In terms of the factor of students' insufficient English abilities, teachers of LAI classes rated much higher ($M = 3.45$, SD = .619) than teachers of AI classes ($M = 3.05$, $SD = .749$) conditions; $t(85) = 2.705$, $p = 0.008$.

Therefore, to present findings of the influential factors, this study focuses on the two evident aspects emerging from the survey results. One is the washback effects, including the two factors of the high-stakes and the low-stakes exams. The other one is regarding students' inadequate skills in English. Then, other notable factors, evidence for which was found quantitatively rather than qualitatively, are presented at the end of this section.

Table 15 *Difference between the two groups*

Influential factors	Group	*n*	*Mean*	*SD*	*Std. Error Mean*
Curriculum (the Course of Study)	LAI	47	2.47	.718	.105
	AI	40	2.60	.709	.112
University entrance exams	LAI	47	2.70	1.041	.152
	AI	40	3.38	.740	.117
Term exams / Regional mock exams (*moshi*)	LAI	47	3.13	.711	.104
	AI	40	3.15	.700	.111
Pre-teacher training experiences	LAI	47	2.13	.711	.104
	AI	40	1.98	.733	.116
On-going teacher training experiences	LAI	47	2.21	.720	.105
	AI	40	2.30	.608	.096
Students' insufficient speaking skills in English	LAI	47	3.45	.619	.090
	AI	40	3.05	.749	.118
Teachers' insufficient speaking skills in English	LAI	47	2.91	.803	.117
	AI	40	2.75	.870	.138
Number of students in a classroom	LAI	47	3.00	.909	.133
	AI	40	3.13	.853	.135
Textbooks	LAI	47	2.77	.813	.119
	AI	40	2.75	.927	.147

4.4.1 Washback effects and professionalism.

4.4.1.1 *Questionnaire*.

The most significant finding from the comprehensive data of the teacher survey was the considerable pressure of term exams and *moshi* (regional mock tests) ($M = 3.14$). Despite these exam types being considered low-stakes, the results indicate that the exams held much sway over the course organization. As noted, teachers seemed to use term-exams as a preparation tool for university entrance exams. Even low-stakes exams may be functioning equally as a constraint on implementation of CLT. Although the mean of university entrance exams was reported to be weaker ($M = 3.01$) than the low-stakes exam factor, the factor had the largest number of "strongly agreed" responses (41%) among the listed factors.

The results of the t-test answered this statistical gap; that is, teachers in AI classes valued the university entrance exam factor as a strong influence ($M = 3.38$, $SD = .740$) higher than ones in LAI classes ($M = 2.70$, $SD = 1.041$) conditions; $t(85) = 3.417$, $p = .001$. This is an understandable result, considering all the aforementioned data suggested that the high-stakes exams seemed to be given strong importance in classrooms. The data indicate that although term exams had influence on both types of classrooms, in terms of university entrance exams, teachers in LAI environments were less affected by the factor. Hence, the data statistically confirmed that negative washback effects existed in this context.

4.4.1.2 *Interviews*.

Consistent with the quantitative data presented above, the university entrance exams were referred to as the most dominant factor in course organization by all teachers in charge of AI classes. In their opinion, due to the pressure of preparing their students for the exams, their lessons should be practical rather than ideal. Interactive activities were likely to be 'saved' for a

visit of assistant language teachers (ALT) in classrooms. ALT are non-Japanese instructors from

an English speaking country, employed by MEXT through the JET (Japan Exchange and

Teaching) Program to assist EFL teachers in public schools. This is part of the educational policy

change to promote the use of CLT from 1987. When an ALT joins a lesson, the organization of

the classroom teaching seemed to be altered into a more communicative one. Nonetheless,

interviewees mentioned that those visits were just an exceptional, and not welcomed, opportunity,

especially in AI schools. Teachers working for AI classes stressed that they did not have any

single second to waste as they prepared their students for the final goal, which is the university

entrance exams.

The following excerpt is from the interview with Ms. Inoue when she disclosed that

teachers were, in point of fact, intentionally ignoring the course objectives of OC courses in the

school:

えーと、あの時間的に、あの、文法的なこともね、オーラルの時間帯にやってるんです。

だから…。言ってしまうんですけど、週３時間がオーラルなんですけど、あの、１時間

はまぁ効く話すみたいな授業で、活動の時間。あとの２時間が文法中心なんです。だか

ら、こう、はい。

[Translation: Well, time-wise, ah, we also teach grammar in lessons of Oral

Communication. So… To be candid, among three lesson hours a week for Oral

Communication, ah, we use an hour for listening and speaking lesson, which means time

for communicative activities. Then, the other two hours are grammar-focused lecture

time. So, you know.] (Ms. Inoue, Interview: 05/26/2010)

Regarding term exams, Ms. Inoue stated that she designed the exams with materials she used in "the other two hours" that were devoted to grammar exercises. Ms. Toyama also avowed that teachers' ultimate aim had to be preparation for the high-stakes exams, and that those who blindly stuck to CLT should leave their jobs. In the following excerpt, she illustrates her professionalism:

> お金を貰ってやっている以上、生徒がメインであるから、生徒が求めるものをやってい
> かないといけないと思うので …(中略) … 自分がやりたいのはこれだからってやるのは違
> うと思うんです。それは自分のエゴだと思うんです。それがやりたいんだったら自分で
> 英会話教室でも開くなり別の形でやらないといけない。この県の教員としてここにいる
> 以上は、学校のニーズ、受験で国公立に行かしてくれっていうなら、それが出来る教員
> じゃないとプロじゃないんだろうなって。
>
> [Translation: As far as we are paid, because the main concern should be our students and their profit, I believe that we have to do things that our students need… So, I think it's wrong if one does whatever he wants to do in a classroom. That's just egoism. If you want to do [CLT], you have to do it in another way such as opening your own English conversation school. As long as you are an employed teacher of this prefecture, you have to meet your school's needs. If they ask you to make your students pass national or public universities' exams, you have to be capable to do so. Otherwise, you are not professional.] (Ms. Toyama, Interview: 05/19/2010)

A teacher in a LAI school whose students would not write the Center Test, Ms. Kurano, also affirmed that she would need to change her teaching practice depending on the school at which

she worked. Although the interviewees were aware that developing the students' speaking skills was the course aim, they held a strong belief that the status quo would continue without a drastic change to the whole university entrance exam system.

4.4.2 Belief in students' incapability in English.

4.4.2.1 *Questionnaire*.

The comprehensive results of the teacher survey revealed that the teachers regarded students' speaking skills ($M = 3.26$) as the most influential obstacles on their course organization. The majority, 88%, of the teachers agreed that the students' inadequate ability restricted them from setting up interactive tasks.

Moreover, the results of the t-test disclosed that a significant difference between the two groups existed in this aspect as well. Teachers of LAI classes rated much higher ($M = 3.45$, SD = .619) than teachers of AI classes ($M = 3.05$, $SD = .749$) conditions; $t(85) = 2.705$, $p = 0.008$. The result is of importance as teachers in LAI classes, although they had much lower negative washback effects than teachers in AI classes, still did have a different reason to avoid CLT in their classrooms.

4.4.2.2 *Interview*.

The interviewees from LAI classes confessed their rationale of why they did not adapt CLTT by referring to this factor. Not only they but also teachers of AI classes, unexpectedly, claimed that their students were not sufficiently proficient to be taught and tested in a communicative manner.

Regarding CLT, in general, the idea of using the target language to create a more communicative language learning environment was highly criticized by the interviewees, since they held a strong view that students would be less interested in learning English. For instance, a

teacher working in a LAI school revealed her uncertainty about the use of English as a medium of instruction. She said that students, many of whom could not understand junior high school level English, would get distracted if she used English to communicate with them. According to her, communicative language teaching would create confusion for students, and that communicative language testing would frighten them; therefore, it was better to abstain from CLTT.

As the survey results show, teachers in AI schools also shared this apprehension. Mr. Nishida, although teaching in a highly AI school, was also concerned about his students' abilities in English, especially their listening skills:

まぁ全部英語でって言うのはまず不可能ですよね。英語を…学んでいる言語を学んでいる言語で教えられても、分からないところが生徒に出てきて、そして途中で付いて来れないようになったら、シャットアウトしてしまうと思うので。そして聞かなくなるだろうなって不安がありますので。

[Translation: Well, it's definitely impossible to conduct a lesson all in English. If we try to teach students English through the exact language they are learning, it would just beyond them. Then, once they know it's hard to follow my instruction, they would be uninterested and inattentive. I worry about that.] (Mr. Nishida, Interview: 05/17/2010)

The interviewees considered that conducting lessons exclusively in English would harm the students' "positive attitude toward communication" due to their inability to understand the instructions.

4.4.3 Other factors: The Course of Study and teachers' own insufficiency of English.

4.4.3.1 *Questionnaire*.

The results of the t-test did not show any statistical differences between the two groups among the other listed factors. However, some weak factors should also be noted to discuss its insignificance in the EFL context.

Firstly, the curriculum was not considered as a strong influential factor ($M = 2.53$), which was also evident from the classroom observations and the samples of term exams. Secondly, it should be noted that training experiences, both pre- and in-service, ranked at the lowest ($M = 2.06$) and the second lowest ($M = 2.25$) by teachers. Lastly, 72% of participants acknowledged that their own speaking skills in English were not sufficient, which may be relevant to the aforementioned interview findings: teachers' profound anxiety over assessing students' speaking skills.

4.4.3.2 *Interviews*.

As described earlier, the teachers who participated in the interviews stated that they did not feel a strong influence from the national guidelines so much to organize their lessons. "It's an ideal goal but it does not fit the reality we are facing." Some hesitantly touched on the fact that refusing to adopt CLT would not result in any penalty to their careers.

Secondly, regarding teacher training experiences, some professional training opportunities were provided by the government and the local school board, some mandatory, but most optional. Even though such opportunities were provided occasionally, considering their regular work would be accumulating day by day, the teachers showed reluctance to devote their time to those opportunities. Mr. Nishida expressed his frustration with the heavy workloads. His time tended to be engrossed by extra tasks besides teaching such as administrative tasks and

extracurricular club activities; thus, he could rarely create time to participate in training opportunities. Likewise, Ms. Hori explained that such heavy workload on diligent EFL teachers was likely to hinder them from getting a job promotion in the current high school system. Moreover, even if they could afford the time for participating in teacher training, the quality of the training seemed to be very questionable.

Lastly, along with the students' incapability of understanding English, some teachers confessed that they were not inclined to use the CLT approach owing to a lack of confidence in their own English. Mr. Abe stated that he sometimes thought his negative attitude towards communicative English might discourage his students from speaking English. He admitted that he himself did not enjoy speaking English, implying that his confidence level of speaking the language was not high. Consequently, as presented in the following excerpt, some of his students started shying away from communicative activities:

…あまり僕自身が英語を使おうとしてないのかなって。えーと、なんていったらいいんでしょうね。たとえば、学習とした初期の段階で、生徒で英語っぽい発音で言おうとする子がいるじゃないですか。そういう子が回りに笑われるような雰囲気を僕は作ってしまっているとぼくは思うんです。そういう子がコーラスリーディングをしたり、発表したりしたときに、声が出なくなってしまってる…という風に追い込んでしまっている気がするんですよね。

[Translation: …I guess I myself am not trying to use English in my classroom. Well, how should I put this? For instance, at the first stage of their learning experience, some students may try to pronounce words in an English-like way, right? Those kids tend to be laughed at by other students. I think it's me who creates this kind of negative atmosphere.

I feel like I am putting them in such a difficult situation. As a consequence, those kids

avoid pronouncing English when they do chorus-reading or a presentation.] (Mr. Abe,

Interview: 05/13/2010)

Although the teachers' insufficient English abilities were not ranked highly in the survey, this

aspect may be in relation to their refusal of CLTT. Only a few interviewees unveiled their own

English competency as a hindrance of OC courses; however, some other interviewees touched on

this not-openly-discussed issue by referring to the forthcoming Course of Study in 2013. The

forthcoming educational policy aims to promote the use of English as an instructional medium.

"I wonder how some teachers cope with the change," said Ms. Kurano. She implied that teachers

who were not fluent in English would have difficulties in conducting lessons. On the other hand,

this change also means the end of OC courses, integrating the courses' core principles with the

new "more comprehensive" courses. "This is a very favourable move for those of us who have to

prepare students for the exams"; in Mr. Nishida's view, the upcoming Course of Study would

bring contradictory effects in high schools. The elimination of OC courses from the curriculum

would simply allow teachers to stay in the traditional grammar-translation methods. In fact, those

who seemed less confident in their English proficiency level mentioned that they would not use

English but would "stick to Japanese".

4.5 Summary

In the above sections, I have discussed the findings related to the two research questions

that deal with the central issue—the curriculum-teaching-testing congruence. In the following

section, I present a summary of the results.

4.5.1 Teaching and testing out of alignment.

The first research question investigated in this thesis concerned how teaching and testing in OC courses were related with each other when the educational policy and the university entrance exam system were incongruent. At various points throughout this study, a mismatch was observed between the course objectives and classroom practices. The results show that while the tenets of the educational policy were adapted to teaching to some extent, testing seldom reflected the policy, indicating that the influence of university entrance exams was more evident in it. Although the goal of the OC courses was supposed to be to help students achieve a fair degree of communicative competence in English, the term exams were not created to share the same goal by clearly lacking the component of measuring students' oral production skills. The interviewees openly admitted that the Course of Study was not in their minds when assessing students. Therefore, teaching and testing in the speaking-focused courses did not show any coherence with the objective of the educational policy in this context.

4.5.2 Negative washback and other influences.

The second research question was also related to the central issue, but the focus was on teachers' perceptions of constraints on OC courses. In line with the interview data, the survey results also strongly indicate that the Course of Study had a very weak influence on teachers. On the other hand, both low- and high-stakes exams were reported as an important factor of teachers' decision making in OC lessons. Particularly, most teachers in AI classes "strongly agreed" that the university entrance examinations controlled their classroom practices. Furthermore, the factor of students' insufficient English skills was rated highly by the teachers of LAI classes as the most considerable influence. Qualitative data also revealed that teachers' self-

reported deficient English speaking skills turned them away from implementing the educational policy.

In this chapter, I have analyzed the data in detail, including the questionnaire data and the qualitative data from classroom observations, term exams, and interviews. The analysis of all the data helps to demystify the uncertainty in the literature around the topic of washback effect in EFL in Japan. Some of these data have provided answers to the research questions. The findings of this research may help to shed more light on this area of debate which will be discussed in the following chapter.

Chapter 5: Discussion of Results

5.1 Introduction

This chapter begins by discussing the main results in relation to certain background variables and related research in the area. The discussion broadly follows the research questions and builds on the analyses of the findings presented in Chapter 4.

Consistent with the findings of the literature, this study also revealed that the teacher factor played an important role in the mechanism of negative washback. The analysis of the four data sets indicated that the teachers' individual characteristics emerged as influential variables, which also echoes the washback literature. In light of the theoretical framework discussed in Chapter 2, this chapter argues the significance of the teacher factor in washback within the scope of foreign language education. Therefore, presenting the salient results that address the research questions, it discusses the global theme of this study—the teacher factor in washback—within the following three categories. The first category covers the teaching practices in relation to the curriculum-teaching incongruence. Then, the second category pertains to the classroom testing/assessment practices. Lastly, the teachers' perceptions on influential factors are discussed, which mainly revolve around, but are not limited to, the issue of washback. Other influential factors that are interwoven with the washback phenomenon in this context are included in this category.

5.2 Teaching in OC courses: Disguised Classroom Teaching

In terms of teaching practice, two issues emerged from the findings: the absence of national guidelines and teachers' instructional use of languages, both of which suggest that the teachers' central concern was not the development of students' speaking abilities, but their preparation for university entrance exams.

5.2.1 The trace of the policy in teaching practices.

The goal of the OC courses is to promote students' communicative competence in English, and, as presented in the previous chapter, the data collected in this study showed some attempts by teachers' to incorporate aural/oral activities into their classrooms. This result may be evidence that the persistent promotion of CLT by the government bore some fruits in classrooms. However, a close examination of the classroom observations and interviews revealed that these activities were not designed to achieve the course aims but rather to disguise the fact that the grammar-focused activities were still predominantly used. This finding echoes the literature that describes the unrelenting dominance of grammar-translation instruction in EFL classrooms in Japan (Ichikawa, 2005; Kikuchi & Browne, 2009; Sato, 2003; Taguchi, 2005; Watanabe, 1996).

As noted earlier, the teachers who volunteered to take part in the classroom observations were assiduous EFL educators, recommended by the Board of Education; therefore, it is legitimate to speculate that their teaching practices were emblematic, or even pioneering, of the high school EFL teachers in the prefecture. Nevertheless, they displayed uncertainty about the definition of communicative activities and CLT methodologies. For example, they labelled activities that required students to pronounce English, even only a single word, as speaking activities, such as chorus reading, recitation, and read-aloud practice. The activities were organized in the form of autonomous pair/group work, which seldom accompanied teacher feedback. Questionably, those mechanical practices were largely coupled with Japanese translation exercises, but not with meaning-focused activities that might effectively elicit learners' pushed output (Nation & Newton, 2009). This finding is in line with Taguchi and Naganuma's (2006) classroom observations where students' English utterances were merely choral repeating of the dialogues. Moreover, the interview data revealed that the chief purpose of

the 'speaking' activities was to create a slightly trendy atmosphere that distinguished it from other conventional grammar-focused courses, but not more than that.

Regarding the importance of interactive activities, Swain (2005) argues that pushed output helps learners attain/improve their L2 as they can transfer their receptive knowledge of the L2 into productive use. In the same vein, Lightbown and Spada (2006) claim that teachers should value communicative tasks in CLT environments rather than focusing only on grammatical analyses. Lacking a provision of those language learning opportunities, the classroom practices did not match the goals established by MEXT.

Moreover, listening, which was reported as the most popular activity in the questionnaire, also implies that teachers' real intention was not to develop students' listening skills but to prepare them for the university entrance exams. The listening tasks in the observed classrooms did not provide the students with an interactive environment, but left them to individually work on multiple-choice questions. This type of listening activity is considered "one-way listening" according to Nation and Newton (2009). They claim that this approach fails "to capture the richness and dynamics of listening as it occurs in our everyday interactions" (Nation & Newton, 2009). Again, based on the observed classrooms, it is possible to speculate that this one-way listening approach may be predominant in other EFL classrooms as well. Considering the fact that the listening component in the Center Test is in a format of multiple-choice questions, test-preparation seems the chief purpose of the popularity of listening activities. Based on this finding, it is possible to assert that the lack of test validity led the teachers to focus on teaching test-taking strategies (Messick, 1989). Furthermore, as Wall and Horák (2006) argued, the insufficient teacher training (on CLT in this context) resulted in a lack of awareness about the change of the test.

From the results of this study, although teachers were showing a hint of shifting from the traditional grammar-translation method to a more communicative method, the overall organization of OC courses still leaves much to be desired to achieve the national guidelines. This updated reality of OC courses also echoes the literature that concludes the MEXT's continuous attempts to promote the use of CLT have failed to be accepted at the classroom level (Gorsuch, 2000; Hiramatsu, 2004; Kikuchi & Brown, 2009; Taguchi, 2005; Taguchi & Naganuma, 2006).

5.2.2 The comfortably status-quo teaching practices.

Next, the teachers' use of language appeared incongruent with the theoretical principle of the CLT approach through the data. Whereas limited attempts to integrate a quasi-CLT approach were found, the central tenet of their pedagogical philosophy still lingered on grammar-translation methodology. Japanese was used as the main medium of instruction by teachers in order to explain the grammatical, especially syntactical, features of the target language. This finding is in accordance with Taguchi's (2005) study which reveals that 93% of the teachers reported using Japanese as the instructional medium to conduct their OC courses. My interviewees emphatically averred that Japanese played an important role in classroom management, presenting rules that govern grammar, and checking for students' comprehension. Nonetheless, my observations of the four classes and the interviews with nine teachers pose a further question. Was the use of Japanese effective in facilitating students' language acquisition? The teachers' justifications of the instructional use of Japanese are open to discussion due to the following two points; 1) L1 is overused; 2) the purpose of the L1 use is ineffective.

In fact, the advantageous use of L1 in the EFL classroom has been argued by scholars in second language acquisition (Auerbuch, 1993; Nunan & Lamb, 1996). However, the contexts of

the studies that acknowledge the positive role of the mother tongue in the classroom are meaning-focused, communicative language learning environments. Tang (2002) suggests that the chief medium of communication in EFL classrooms should still be English, although "limited and judicious use" of L1 can aid learners' language acquisition processes. She concludes that no more than 10 percent of class time should be spent using the L1. As in this study only 3% of the teachers were using English as the main language in the speaking-focused classrooms, the nature of the use of L1 in the investigated context was fundamentally different from those in the literature. The OC courses were conducted in a form-focused, non-interactive, conventional teacher-centered environment, where Japanese was overused and English played only an ancillary role.

Secondly, based on the findings of the classroom observations, it was evident that the use of L1 was not carefully planned to facilitate students' understandings of English. Teachers used the L1 frequently to translate their own utterances in English, even for simple sentences. This tendency might automatically sway students away from paying attention to teachers' instructions or messages conveyed through the target language. The teacher-student interactions, moreover, were taking place in the L1, not only in language-focused tasks but also in the aforementioned aural/oral tasks. In his Interaction Hypothesis (1996), Long advances that the opportunity for pushed output through comprehensible interaction is conducive to learners' language development. Thus, in this sense, the teachers are likely to squander the chance of providing effective learning experiences for their students.

Therefore, based on the aforementioned theories of CLT, it is plausible to interpret that significant amounts of the use of Japanese were generated by the needs to prepare students for the university entrance exams, especially in AI classes. Without a fundamental change in the

high-stakes testing system, it seems that teachers can comfortably maintain the same conventional teaching environment. As shown in Pan's (2009) model of negative washback, the high-stakes exams led teachers in this context to narrow the curriculum and lose instructional time, and teach to the test.

5.3 Testing in OC courses

Turner (in press) argued that classroom-based assessment (CBA) is "a contextually-bounded and socially-constructed activity involving different stake-holders in learning" (p. 3). The findings of the study can be an example of the nature of CBA—testing/assessment managed by the classroom teacher to serve students' learning—in relation to washback. As discussed in Chapter 2, Canale and Swain's (1980) model of Communicative Language Testing proposes that learners have to be tested not only on their knowledge of language, but also on their ability to put it to use in a communicative situation. In this study's context, however, the findings confirmed little trace of the principle in classroom testing, which supports Akiyama (2004) and Taguchi's (2005) arguments. They questioned the quality of classroom testing that did not reflect the course aims under the strong influence from university entrance exams. In this study, more evident than the teaching practices, the intention to develop students' speaking skills was missing in the three following aspects of the testing practices. Firstly, the assessment tools—term exams—were misused to prepare students for high-stakes exams. Secondly, teachers' misconceptions about the methods of assessing speaking were found. Lastly, it was found that the need to discuss teachers' negative attitudes towards speaking tests stemming from their uncertainty, or even aversion, about assessment.

5.3.1 Misapplied assessment tools under the washback effects.

The findings point to dark shades, negative influences, of university entrance exams in the low-stakes exams. Echoing Akiyama's (2004) study, the findings suggest that the major internal testing tools, school-based term exams, were administered in a paper-pencil fashion even in the speaking courses. Therefore, for the sake of practicality, the term exams excluded speaking tasks but significantly included language-focused tasks to test grammatical knowledge, writing skills, and translation skills. It is also observed that besides the official textbooks, sub-materials for grammar drills and vocabulary building seemed to be commonly used to design the tests, especially in AI schools. Those additional materials were not used in teaching but included in testing. This particular finding indicates that the negative washback effects were more evident in classroom testing than teaching. Thus, term exams were used as tools to transmit a subliminal message that students do not need to hone their speaking skills to pass university entrance exams and the high-stakes exams are the ultimate goal to achieve. In this context, the university entrance exams had the strong influence on the teachers' CBA practices negatively, resulting in students not being "able to learn real-life knowledge, but instead learn discrete points of knowledge that are tested" (Pan, 2009).

Clearly, this situation questions the organization of all the OC courses in the absence of an important part of students' learning experience. According to Nation and Newton (2009), to obtain reliable information to measure learners' progress, it is important to observe them "while they are involved in listening and speaking activities" (p. 165). Interactiveness is one of the criteria test developers need to consider, especially when the tester has to obtain the information of a test-taker's productive skills (Bachman & Palmer, 1996). The paper-based term exams used in the target context cannot possess such characteristics. Morrow (1981) pointed out that

communicative tests should be assessed qualitatively rather than quantitatively, which, again, did not match the nature of the presented term exams in the OC courses. Therefore, teacher-testers cannot allege that the term-exams are relevant for use in the specific courses.

In accordance with the finding that CLT was not happening in the classrooms, the testing practice also leaves much to be improved in this context.

5.3.2 Myths of assessing speaking skills.

As a consequence of the long-lasting washback phenomenon in this context, the lack of urgency to test/assess students' speaking skills seems to hinder teachers from developing assessment literacy for speaking. This finding echoes the literature that discusses washback on teacher perceptions (see e.g., Cheng, 1997, 2000; Turner, 2001, 2006, 2009; Wang, 2010; Watanabe, 1996), and to be specific, it provided some empirical evidence of the test impact on teacher perceptions of CBA.

As discussed above, while most of the teachers did not provide any speaking tests to students, a small number of teachers, albeit not frequently, administered tests to measure their students' speaking abilities. Among the very few teachers, there were prevalent myths about assessing oral production. That is, their grading scales were designed to quantify the candidates' kinetic performance rather than constructs of oral production such as fluency, accuracy, and content. When it came to speaking tests, teachers turned into loyal supporters of the Course of Study, referring to a goal of the courses that aims at "fostering students' positive attitudes towards English." Thus, they used the grading rubrics in which kinetic messages were interpreted as English competency. Quantifying the test-takers' attitudes tends to be highly subjective, and, in the first place, tests that do not focus on learners' English proficiency should not be used to gather information about their learning progress (Underhill, 1987).

Another myth was that students' speaking abilities could be assessed from the quality of non-extemporaneous production, such as written essays, reading comprehension, and even recitation tasks. Skehan (1990), in particular in relation to CLTT, points out that although there always are limitations in direct testing, indirect language testing has the tendency of failing to capture valid information to appropriately assess test-takers' communicative competence. Therefore, if the target language domain of the test is speaking, then it is of great importance to employ direct testing using performance tasks.

In short, this specific finding of teachers' misconstructions illustrates their lack of assessment literacy for speaking, which may stem from the serious dearth of teacher training on CBA as well as the washback effects.

5.3.3 Misgivings about communicative language testing.

In line with the two aforementioned findings, one more area that should be further discussed is the severe scarcity of teacher training opportunities and the allocation of resources to enable teachers to participate in them. Training experiences is an imperative characteristic of the teacher factor (Turner, 2006, 2009; Wang, 2010; Watanabe, 1996, 2000).

In this study, none of the interviewees had had a chance to receive any in-service training for assessment, which could exacerbate the incongruence between teaching and testing practices. As is argued by Bailey et al. (1996), without reasonable training, teachers are prone to fall back on how they themselves were taught as students, a phenomenon known as 'the apprenticeship of teaching' (Kikuchi, 2006). Thus, the dearth of teacher training for testing, and for speaking in particular, might also have been a contributing factor generating the deep-rooted anxiety for teachers. Shohamy (2000) argues that the two areas of language testing and second language acquisition should cooperate for the sake of learners' beneficial language learning experience.

Along with the need for sufficient training and on-going support for teachers, stakeholders should verify that not only classroom teaching but also the testing conforms to the course aims.

5.4 Influential Factors of OC courses

As discussed earlier, the testing and teaching practices in the OC courses did not match the course objectives, missing the most crucial domain of the aural/oral-focused course. In this section, some significant issues of influences emerging from the data set are discussed. Along with the main issue of the study—negative washback—other influential factors that may interweave with it are addressed. Some of the results in this study depart from those of the literature, which could be the result of the special circumstances of the investigated demographics.

5.4.1 Washback: What is the real goal for EFL teachers?

A significant finding from the survey was that the low-stakes exams also had a strong impact on teachers as well as the university entrance exams in the speaking-focused courses. As noted, even in the case of low-stakes term exams, teachers used them as preparation for the high-stakes exams; therefore, echoing the literature, this can be interpreted as washback effects on CBA (Akiyama, 2004; Taguchi, 2005). This point is further supported by the results from the paired sample t-test, which revealed that teachers in AI classrooms reported the influences of the university entrance exams as stronger than ones in LAI classrooms. The qualitative data were in line with these findings as teachers, especially those who were working in AI environments, referred to the impact of university entrance exams to justify their classroom practices.

Moreover, the qualitative data suggested that the university entrance exam had a significant influence on EFL teachers' perception of their role. The professionalism was described in relation to the high acceptance rate of renowned universities, not students' linguistic

development. Therefore, this finding can also be evidence of negative washback effects over language teachers (Alderson et al., 1990; Cheng, 1997, 2000; Shohamy, 2004; Watanabe, 1996).

Described in the literature, the washback effect relates to the contextual variables of the society in which the test is used (Alderson & Wall, 1993; Bachman & Palmer; 1996). Owing to the magnitude of importance entrance exams possess in Japanese society (Buck, 1988; Guest, 2008; Reesor, 2002), teachers feel the influence of the high-stakes exams not only directly but also indirectly from society, such as the expectation from students, their parents, the community, and the school. This echoes the theoretical statement of Shohamy et al. (1996); that is, if the stakes of a test are high in the society, its influence over its stake-holders will also be strong.

5.4.2 Students' insufficient English ability.

As the most influential factor, students' deficient English competency was recognized by the teachers, which corresponded to Gorsuch's model (2000). Although the results of the t-test revealed that teachers in LAI classrooms rated higher than ones in AI classrooms on this factor, both groups considered it as a significant obstacle.

In the case of LAI classrooms, the quantitative data indicate that this factor, the students' low English proficiency, was the main challenge for the teachers to move away from the traditional, grammar-focused, and teacher-centered, teaching style. In line with the above point, in the qualitative part, the teachers deemed that adaptation of CLTT would do more harm than good owing to students' incompetency in English. Likewise, teachers in AI classrooms held the view that a communicative learning environment was inappropriate for their students. In Gorsuch's argument (1998), even teachers at top-ranked schools avoided interactive activities, due to their conviction that those activities would be "too difficult" for their students, which applies to this context as well.

5.4.3 Rejected policy: Tacit approval of the hidden curriculum.

As noted earlier, as a highly controversial fact, other unofficial sub-textbooks were largely used in classrooms to prepare students for the university entrance exams. As a consequence, the participants in this study did not feel the strong influence of the official textbooks authorized by MEXT. Especially in term exams, those sub-materials were used and given more emphasis than the official ones. Due to socio-cultural norms and values, "the hidden curriculum" exists as a major influence on what teachers accept as usual and normal in their classrooms (Denscombe, 1982). Under such a circumstance where teachers were allowed to ignore the policy, the textbook designed in conjunction with the policy had little influence in their classrooms. The weak impact of the Course of Study has been the topic of heated discussion (Gorsuch, 2000; Nishino & Watanabe, 2008). The current policy does not serve its own purpose and its own aims but only exists as an unattainable dream.

One important finding in this study that should be further investigated is that the participants had already foreseen that the forthcoming Course of Study in 2013 would not be backed up by teachers disagreeing with the educational change in which teachers are supposed to use English as a medium of instruction. Considering the fact that the current policy failed to be adapted into EFL classrooms, without a radical change in university admission systems, it is highly possible to speculate that teachers will continue with the familiar teaching style.

In O'Donnell's (2005) findings, the teachers in the study wished that in tandem with the Course of Study, entrance exams for universities included a speaking component for positive washback. Even though it might not be so straightforward to create positive washback in all cases, it is essential to consider potentially relevant factors that researchers recommended in

order to create a test with positive impact. Hughes (1989) claims that the teacher factor is of

great importance in accord with the introduction of a communicative language test:

> "One important reason for introducing the new test may have been to encourage
>
> communicative language teaching, but if the teachers need guidance and possibly training,
>
> and these are not given, the test will not achieve its intended effect" (p. 46).

From this point of view, along with a discussion of the inclusion of a speaking component in

entrance exams, it is crucial to consider what factors may aid teachers in facilitating positive

washback.

5.4.4 Teachers' profound anxiety as non-native English speakers.

The last issue that needs to be addressed is the teachers' lack of confidence to employ the

CLTT approach. Although the factor was not salient quantitatively, the factor of teachers' lack of

confidence in oral competency qualitatively emerged, corresponding to their aversion to

assessing speaking abilities of the target language. This characteristic of the teacher factor is a

crucial problem in EFL contexts that may accelerate the negative washback effects (Wang, 2010).

Teachers' deficiencies in oral English have been discussed in the literature (Hiramatsu,

2004: Sato, 2002). Hiramatsu (2004) documented how the teachers' insufficient English

proficiency was obstructing them from adapting CLT into their classrooms, not being able to

cooperate with their ALT. This phenomenon may be connected to the profound anxiety over

communicative English that Japanese EFL teachers hold as non-native English speaking (NNES)

teachers (Samimy & Brutt-Griffler, 1999; Hiramastu, 2004). The argument revolves around

NNES teachers' concerns over their own linguistic inferiorities, which tend to demotivate them

from communicative instruction. As Richards and Lockhart (1994) contend, teachers' beliefs and

self-perceptions often influence their teaching behaviour. Hence, it is not surprising that teachers

tended to shy away from engaging their students in speaking activities. Even in cases where some interactive activities were actually organized, they did not venture into setting up testing opportunities. This seemed to be due to their uncertainty about evaluating communicative competence, which required them to also have solid English language skills.

Even though this issue has been discussed *ad nauseam* in Japan, we have not seen any fundamental change yet. One reason behind this enduring status-quo may be attributed to the educational culture in Japan. Hiramatsu (2004) proposes a culturally embedded explanation. Making mistakes in meaning-focused activities, she suggests, could be a humiliation for Japanese EFL teachers in a culture in which teachers and other leaders "should know everything." The teachers' imperfection, therefore, is likely to be concealed as a taboo. However, as the successful implementation of the forthcoming new policy depends on teachers' communicative competency in English, and most importantly, for the sake of students' better learning experience, urgent needs of teacher training that are specifically designed for NNES teachers should not be left out from the discussion.

5.5 Summary

In this chapter, I interpreted and discussed the findings regarding the OC courses. Even though the courses had been running for two decades, the updated data suggest that the reality of the current classrooms largely remained the same as reported in the literature. Key issues regarding the research questions are; 1) the curriculum—teaching-testing incongruence, and 2) the tremendous influence of the university entrance exams on teacher perceptions of the speaking-focused courses.

The most prominent finding was the obvious curriculum-teaching-testing incongruence of the OC courses. Without providing an adequate support system for EFL teachers to develop their awareness of this issue, the stagnant situation is bound to continue.

The other key issue concerns teachers' perceptions of the negative washback effects. Although further studies should be conducted in this context to investigate the complex nature of washback, it was apparent that teachers valued the high-stakes exams more than the development of students' speaking skills in the context where the educational policy and the university entrance exams are out of alignment. Thus, in tandem with a fundamental change of the university entrance exam system, especially the Center Text, MEXT should consider whether the policy sets attainable goals and fits the reality of high school EFL education.

Chapter 6: Conclusions and Implications

6.1 Introduction

This chapter discusses tentative conclusions and some suggestions for further research. It starts with the outline of the study. Then, it addresses the theoretical and methodological limitations, and it concludes with implications for EFL classroom practice, and especially testing, of EFL speaking in high school contexts.

6.2 Research Summary

In light of Japanese language policy statements and the national curricular goals aiming at the development of students' practical communication abilities, the purpose of the study was to examine how EFL teachers in high schools perceived their teaching, testing, and influences on speaking-focused courses, both at general and specific level. In this study, it was evident that high-stakes exams had a strong impact on classroom practice, which supports the claims in the literature on negative washback (see, e.g., Bailey, 1996; Cheng, 1997; Lam, 1994; Wall & Alderson, 1993). Moreover, in terms of professionalism, it seemed to even metamorphose out of language teachers into test strategy instructors. Other constraints that were distinctively found in the research context appeared to amplify the negative washback effects.

Firstly, in teaching practice, some attempts to accommodate the national policy were observed. However, over all, the classroom environment of OC courses still remained teacher-centered, grammar-oriented, and mono-lingual, where listening and speaking activities were organized in a one-way fashion. This finding is in line with that of the literature (Ichikawa, 2005; Kikuchi & Browne, 2009; Sato, 2003; Taguchi, 2005). The qualitative findings suggested that the teachers' main concern was to design "OC-ish" activities that would help students earn high scores on the university entrance exams. In accordance with Taguchi's (2005) findings, Japanese

was predominantly used to interact with students in the classrooms. In this study, although the two teacher groups held different reasons for the use of the L1, Japanese was largely used as a medium of instruction regardless of the students' academic levels; thus, it was evident that the CLT approach had not been adopted at the local public high school level yet.

Secondly, negative washback effects were more evident in the assessment of the courses than in teaching, indicating that classroom teaching and testing practices were not congruent with each other or with the course objectives. The absence of a clear guideline for assessment and of professional development opportunities emerged as a key factor of the incongruence. Without question, teachers were misusing the term exams as a preparation tool for university entrance exams. Even in the speaking-focused courses, most teachers reported that they graded their students without testing their oral production, which is consistent with other findings in the research literature (Akiyama, 2004; Taguchi, 2005). A distinctive tendency found in this study was that some teachers actually tested speaking, but their focal point was on students' attitudes and not on the language itself. This finding suggests that under the strong washback effects from the university entrance exams, development of assessment literacy was out of the scope of teachers' interests.

Then, amongst various influential factors, teachers reported that they were highly influenced by exam factors (both high-stakes and low-stakes) more than by the national curriculum guidelines, which echoes the literature (Gorsuch, 2000, 2001; Ichikawa, 2005). The top-down manner of implementing the educational policy that is discordant with the nationwide high-stakes exam, the Center Test, was evidently the cause of this educational disruption. Under such circumstances, contextual constraints such as large class size and time limitation are

perpetuating the status quo. Those constraints can be a reason for teachers to avoid facing insurmountable challenges that require continuous efforts.

As well as the national curriculum, the factors of teacher training experiences were regarded as the weakest influence on their classroom practices. The unsatisfactory quality of the past teacher training and its inaccessibility due to the teachers' heavy workload seemed to discourage them from participating in further professional learning opportunities. Without updating their knowledge and skills to be actively involved in the implementation of the educational policy, EFL teachers tended to keep in step with the mainstream thinking of the Japanese school culture (Kikuchi, 2009). This malfunctioning system, as a consequence, led teachers to attribute the stagnant situation to their students not being sufficiently proficient in English to be taught in a communicative environment. In view of Japanese social norms, it is difficult for teachers to acknowledge their own deficient English speaking competency; however, considering the fact that the factor emerged as an acute problem of the whole issue, this domain should be the center of a teacher training program.

As a final point, chasms exist between the goals of MEXT and the actual classroom practices in EFL classrooms. The new national syllabus will be implemented in a few years, putting more emphasis on the use of English in classrooms. Nevertheless, the educational policy set by MEXT does not appear achievable in the absence of its reflection in actual classroom practices. In order to change the state of affairs, administrators of university exams, teacher educators, and teachers themselves should work together and in harmony to bring about a positive change.

6.3 Implications and Further research

6.3.1 Pedagogical implications.

As implications for teachers, the question that arises in this context is whether test preparation should remain the core of the students' EFL learning experience after two decades of the implementation of OC courses. In this regard, Shohamy's (2004) calls should be heeded in order to draw more attention to teachers' role in language testing. Accordingly, teachers should not be bureaucrats who yield to authority and simply carry out their orders, but should be professionals who are "responsible and involved readers in the field of second language testing" (p. 107). After all, what is most important is that we should not ignore the students' right to be appropriately tested. In accordance with the right of learning, they are entitled to be tested in a manner that enhances their language-learning experience. What is more, even though developing oral proficiency is challenging for NNES, without teachers' "positive attitude towards communication," it would be unfeasible to foster students with such attitude.

6.3.2 Implications for policy makers.

Firstly, the educational policy makers should reconsider to reflect the reality of EFL classrooms and the society of Japan. In view of the fact that the cost of training all JTE would be enormous, Ichikawa's (2006) suggestion seems practical and convincing. He argues that the government should consider providing speaking courses as optional courses so that trained teachers for this purpose can properly serve the students. De Mejía (2002) also claims that bilingual education for elite students can be cost-effective and reasonable.

Secondly, the intrinsic aim of educational policy should be aligned with the nationwide high stakes exam—the Center Test—in order to generate congruence between the policy and the classroom practices. With the inclusion of a speaking component in the test, it is highly possible

to promote the use of CLTT more effectively in this test-driven society, Japan. Using washback

to positively influence language learning has been discussed internationally (e.g., Hughes, 1989;

Bailey, 1996; Saif, 2006; Turner, 2001, 2006, 2009). In light of the points made in the literature

on positive washback, Hughes (1989) suggests the creation of certain particular conditions to

generate positive washback. As a drawback to his proposal, he admits it would be expensive to

execute the ideal model of language testing; however, considering the amount of time and effort

that have been invested in the implementation of the Course of Study, revising its testing system

may be cost-effective in the end.

6.3.3 Research implications.

To date, CBA has been escaping the attention of researchers of the washback

phenomenon as CBA itself as a recognized paradigm is "still in its initial stages" (Turner, in

press); however, perhaps the time has come to bring further attention to this aspect of assessment

to advocate teachers' awareness of "critical testing" (Shohamy, 2004). I hope the results of this

study may instigate and motivate further research.

More work needs to be done on EFL teachers' perspectives on CLTT in different

contexts to see if the results of the study are replicated, or if any recurrent pattern emerges. For

example, it would be valuable to explore to what degree teachers' English proficiency level

impacts upon their performance assessment, and whether it affects students' language acquisition

outcomes. How teachers at different proficiency levels assess students' speaking skills, how

CLTT can be adapted into classrooms by them, and what are challenges for teachers with low

oral proficiency skills, can be questions to be explored. As research on NNES teachers'

assessment skills in EFL speaking is inconclusive in this area, there is much to be investigated.

In the same manner, investigating the effects of CLTT on student motivation and academic

performance would significantly contribute to our understanding of this critical area of research in Japan.

Furthermore, prior to any drastic changes in the university entrance exam system, practical and creative assessment methods for speaking—including tasks and grading rubrics—need to be discussed and carefully evaluated. Such detailed examinations, development processes, and implications will, no doubt, take time. Most importantly, in order for these assessment methods for speaking to be smoothly integrated into this specific context, factors such as time management difficulties, large class sizes, and teachers' aversion toward evaluating English speaking abilities, should be taken into account in the creation of suitable assessment tools.

6.4 Limitations of the Study

The data collection process was an unexpectedly arduous process. At the outset of the data collection, the return rate of the questionnaire was very low, and I called or visited schools to ask for their participation. While most of these teachers stated that they were too busy to answer the long questionnaire, others provided even more telling explanations. Their reason for refusing to participate in the study was that their OC classrooms did not have any problems to be researched. Therefore, unfortunately, the collected data from the questionnaire presented in this study are missing some aspects of the targeted population, even though the refusal explanations may be seen as informative in and of themselves.

Secondly, a weakness of my classroom observations became apparent after the administration. As noted, one teacher's classroom observation data were eliminated due to her disclosing that she had manipulated her lesson to create a more communicative environment for the observation. Since it is possible that the observer's paradox (Labov, 1972, p. 209) may have

affected the other 4 teachers' classroom practices, the classroom observation guide, adapted from Turner's (2000) study, was not fully utilized to present more detailed information of the classrooms as planned. In addition, it is risky to generalize the results of the study, owning to the characteristics of the participants in this segment. They were highly educated elite teachers, recommended by the Board of Education, and in all probability, only representing one type of EFL teachers.

Regarding the test samples, some uncertain points remain. The data imply that there are quite significant differences in terms of test design in each school. In this study, only four sample exams were used, which is insufficient to discuss whether the test design correlated with types of schools, classrooms, or teachers. Likewise, the types of grading scales used by teachers to assess speaking should be investigated further with a larger sample.

Furthermore, the particular characteristics of school teaching culture posed some challenges in the interview component. Having the experience of working in the study context, I was aware that interviewing about teaching practices would be a sensitive topic for teachers, especially about their use of English in classrooms. Whereas most of the interviewees were supportive and provided their candid opinions, some teachers, in fact, expressed some subtle frustration in addressing the subject, which might have created barriers between them and me as the interviewer.

Last, but not least, concerning the data analysis of the questionnaire, I came to realize that the administered t-test is not the best example of the teachers' perspectives. Although the percentage may be not high, LAI classrooms also had students hoping to enter university. Therefore, the participants should be grouped into at least three categories, including a group of

in which none of whose students will write university exams. It would provide more clear insights into the research topic.

6.5 Closing Remarks

In conclusion, by revealing the incongruence of teaching and testing in OC courses, and the clear absence of the educational policy in their implementation, this research highlighted the need for urgent change in the university entrance exam systems that closely reflect the educational policy. In addition, I argued that without proper support for teachers, any change in the educational system would not bring its expected effects in actual classroom practices. It is hoped that these findings will help encourage others acknowledge the role of teacher training for CLTT in foreign language classrooms and stimulate further study in this area.

REFERENCES

Akiyama, T. (2004). *Introducing EFL speaking tests into Japanese senior high school entrance examination*. Unpublished PhD thesis, University of Melbourne.

Alderson, J. C., & Wall, D. (1993). Does washback exist?. *Applied Linguistics*, *14*(2), 115-129.

Alderson, J.C. & Hamp-Lyons, L. (1996). TOEFL preparation courses: a study of washback. *Language Testing*, *13*(3), 280-297.

Anderson, J. O., Muir, W., Bateson, D. J., Blackmore, D., & Rogers, W. T. (1990). *The impact of provincial examinations on education in British Columbia: General report*. Victoria: British Columbia Ministry of Education.

Auerbach, E. (1993). Reexamining English only in the ESL classroom. *TESOL Quarterly*, *27*(1), 9-32.

Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice.* Oxford: Oxford University Press.

Bachman, L. F. & Palmer, A. S. (2010). *Language assessment in practice: Developing and using language assessments in the real world*. Oxford: Oxford University Press.

Bailey, K. M. (1996). Working for washback: a review of the washback concept in language testing. *Language Testing, 13*(3), 257-279.

Bailey, K. M., B. Bergthold, B. Braunstein, N. Jagodzinski Fleischman, M. P. Holbrook, J. Tuman, X. Waissbluth & L. J. Zambo. (1996). The language learners' autobiography: examining the "apprenticeship of observation". In D. Freeman & J. C. Richards (Eds.), *Teacher Learning in Language Teaching* (pp. 11−29). New York: Cambridge University Press.

Bailey, K. (1999). Washback in language testing. *TOEFL monograph series*. Princeton, NJ: Educational Testing Service.

Baker, E. (1991). Alternative assessment and national policy. Paper presented at the National Research Symposium on Limited English Proficient Students' Issues: Focus on Evaluation and Measurement, Washington, DC.

Brown, H. D. (2000). *Principles of Language Learning and Teaching* (4th ed.). New York: Addison Wesley Longman.

Brown, J, D., & Yamashita, S. (1995). English language entrance examinations at Japanese universities: 1993 and 1994. In Brown, J. D., & Yamashita, S. (Eds.), *Language Testing in Japan* (pp. 86-100). Tokyo: Japan Association of Language Teaching.

Buck, G. (1988). Testing listening comprehension in Japanese university entrance examinations. *JALT Journal*, *10*, 12-42.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, *1*(1), 1-47.

Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education*, *11*(1), 38-54.

Cheng, L. (2000). *Washback or backwash: A review of the impact of testing on teaching and learning.* (ERIC Opinion papers ED 442280)

Cheng, L. (2005). *Changing language teaching through language testing: A washback study. Studies in language testing, 21*. Cambridge: Cambridge University Press.

Cook, M. (2009). Factors Inhibiting and Facilitating Japanese Teachers of English in Adopting Communicative Language Teaching Methodologies. *k@ta, 11*(2), 99-116.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage.

Davies, A. (1968). *Language testing symposium: A psycholinguistic approach.* Oxford: Oxford University Press.

de Mejía, A. M. (2002). *Power, prestige, and bilingualism: International perspectives on elite bilingual education*. Clevedon: Multilingual Matters.

Denscombe, M. (1982). The hidden pedagogy and its implications for teacher training: An ecological analysis. *British Journal of Sociology of Education, 3*, 249-265.

Gorsuch, G. (1998). Yakudoku EFL instruction in two Japanese high school classrooms: An exploratory study. *JALT Journal*, *20* (1), 6-32.

Gorsuch, G. (2000). EFL educational policies and educational cultures: Influences on teachers' approval of communicative activities. *TESOL Quarterly*, *34*(4), 675-710.

Gorsuch, G. (2001). Japanese EFL teachers' perceptions of communicative, audiolingual and *yakudoku* activities: The plan versus the reality. *Educational Policy Analysis Archives, 9*(10). Retrieved from http://epaa.asu.edu/epaa/v9n10.html

Hitamatsu, S. (2004). EFL teaching in Japan. In D. J. Tedick (Ed.), *Second Language Teacher Education: International Perspectives* (pp. 113-134). Mahwah, NJ: Lawrence Erlbaum.

Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.

Hughes, A. (1993). *Backwash and TOEFL 2000.* Unpublished manuscript, University of Reading.

Ichikawa, K. (2006). A survey of the actual conditions of "Oral Communication" at high schools, *The journal of Seigakuin University, 18*(3), 239-248.

Kikuchi, K. (2006). Revisiting English entrance examinations at Japanese universities after a decade. *JALT Journal, 28*(1), 77–96.

Kikuchi, K., & Browne, C. (2009). English educational policy for high schools in Japan: Ideals vs. reality. *RELC Journal, 40*(2), 172-191.

Kington, A., Sammons, P., Day, C., & Regan, E. (2011). Stories and Statistics: Describing a Mixed Methods Study of Effective Classroom Practice. *Journal of Mixed Methods Research*, *5*(2), 103-125.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Lightbown, P. M., & Spada, N. (2006). *How languages are learned*. Oxford: Oxford University Press.

LoCastro, V. (1994). Learning strategies and learning environments. *TESOL Quarterly*, *28*(2), 409-414.

Long, M. H. (1996). The role of the linguistic environ- ment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). San Diego, CA: Academic.

Messick, S. (1996). Validity and washback in language testing. *Language Testing*, *13*(3), 241-256.

Ministry of Education, Culture, Sports, Science, and Technology. (2003). *Regarding the establishment of an action plan to cultivate 'Japanese with English abilities'*. Retrieved from http://www.mext.go.jp/enlish/topics/03072801.htm

Moon, T., & Callahan, C. (2001). Curricular modifications, family outreach, and a mentoring program: Impacts on achievement and gifted identification in high risk primary students. *Journal for the Education of the Gifted, 24*, 305-321.

Morrow, K. (1981). Principles of communicative methodology. In K. Johnson & K. Morrow (Eds.), *Communication in the classroom* (pp. 59-69). Harlow, UK: Longman.

Muñoz, A. & Álvarez, M. (2010). Washback of an oral assessment system in the EFL classroom. *Language Testing*, *27*(1), 33-49.

Nishino, T., & Watanabe, M. (2008). Communication-oriented policies versus classroom realities in Japan. *TESOL Quarterly*, *42*(1), 133-138.

Nation, I. S. P. & Newton, J. (2009). *Teaching ESL/EFL listening and speaking*. New York, Routledge.

NCUEE. (2011). *Heisei 23 nendo daigaku nyushi senta shaken jisshi kekka no gaiyou* [Outline of the results of National Center Test in 2011]. Tokyo: National Center for University Entrance Examiantions.

Nunan, D., & Lamb, C. (1996). *The self-directed teacher*. Cambridge: Cambridge University Press.

O'Donnell, K. (2005). Japanese Secondary English Teachers: Negotiation of Educational Roles in the Face of Curricular Reform. *Language, Culture and Curriculum, 18*(3), 300-315.

Oller, J. (1979). *Language tests at school: A pragmatic approach*. London: Longman.

Pan, Y. C. (2009). A review of washback and its pedagogical implications. *VNU Journal of Science, Foreign Languages, 25*, 257-263.

Reesor, M. (2002). The bear and the honeycomb: A history of Japanese English language policy. *NUCB Journal of Language, Culture, and Communication, 4*(1), 42-52.

Richards, J. C., & Lockhart, C. (1994). *Reflective teaching in second language classrooms.* New York: Cambridge University Press.

Saif, S. (2006). Aiming for positive washback: a case study of international teaching assistants. *Language Testing, 23*, 1-34.

Samimy, K. K., & Brutt-Griffler, J. (1999). To be a native or non-native speaker: Perceptions of "non-native students in a graduate TESOL program. In Braine,G (Ed.), *Non-native educators in English language teaching* (pp. 127-144), Mahwah, NJ: Lawrence Erlbaum.

Sato, K. (2002). Practical understandings of CLT and teacher development. In S. J. Savignon (Ed.), *Interpreting communicative language teaching: contexts and concerns in teacher education* (pp. 41-81). New Heven, CT: Yale University Press.

Sato, K., & Takahashi, K. (2003). Teacher and student learning in the workplace: the impact of performance tests. *FALT2000: towards the new millennium. Proceedings of the 26th FALT International Conference on Language Teaching / Learning* (pp. 325-336). Tokyo: JALT.

Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, *13*(3), 298-317.

Shohamy, E. (2000). Using language tests for upgrading knowledge: The phenomenon, source and consequences. *Hong Kong Journal of Applied Linguistics, 5*(1), 1-18.

Shohamy, E. (2004). The power of tests over teachers: The power of teachers over tests. In D. J. Tedick (Ed.), *Second language teacher education* (pp. 101-111). Mahwah, NJ and London: Lawrence Erlbaum.

Skehan, P. (1990). Communicative language testing. *The journal of TESOL France*, *10*(1), 115-127.

Spratt, M. (2005). Washback and the classroom: the implications for teaching and learning of studies of washback from exams, *Language Teaching Research, 9*(1), 5-29.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input and Second Language Acquisition* (pp. 235-253). Rowley, MS: Newbury House.

Swain, M. (2005), The output hypothesis: theory and research. In E Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (pp. 471–483), Mahwah, NJ: Erlbaum.

Taguchi, N. (2005). The communicative approach in Japanese secondary schools: Teachers' perceptions and practice. *The Language Teacher, 29*(3), 3-9.

Taguchi, N., & Naganuma, N. (2006). Transition from Learning English to Learning in English: Students' Perceived Adjustment Difficulties in an English-Medium University in Japan, *Asian EFL Journal, 8*(4), 52-73.

Tang, J. (2002). Using L1 in the English Classroom. *English Teaching Forum, 40*(1), 36-43.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Thousand Oaks, CA: Sage.

Turner, C. E. (2000). Listening to the voices of rating scale developers: Identifying salient features for second language performance assessment. *Canadian Modern Language Review, 56*(4), 555-584.

Turner, C.E. (2001). The need for impact studies of L 2 performance testing and rating: Identifying areas of potential consequences at all levels of the testing cycle. In A. Brown et al. (Eds.), *Experimenting with uncertainty: Language testing essays in honour of Alan Davies* (pp. 138-149). Cambridge: Cambridge University Press.

Turner, C. E. (2006). Professionalism and high-stakes tests: Teachers' perspectives when dealing with educational change introduced through provincial exams. *TESL Canada Journal*, *23*(2), 54-76.

Turner, C. E. (2009). Examining washback in second language education contexts: A high stakes provincial exam and the teacher factor in classroom practice in Quebec secondary schools. *International Journal of Pedagogies and Learning*, *5*, 103-123.

Turner, C. E. (in press). Classroom assessment. In G. Fulcher & F. Davidson (Eds.), *Routledge Handbook of Language Testing.* New York: Routledge, Taylor & Francis.

Underhill, N. (1987). *Testing spoken language: A handbook of oral testing techniques*. Cambridge: Cambridge University Press.

Vernon, P. E. (1956). *The Measurement of Abilities.* (2nd. Ed.) London: University of London Press.

Wang, J. (2010). *A study of the role of the 'teacher factor' in washback*. Unpublished doctoral dissertation, McGill University, Montreal.

Wall, D. and Horák, T. (2007) 'Using baseline studies in the investigation of test impact', *Assessment in Education: Principles, Policy & Practice*, *14*(1), 99-116.

Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing*, *13*(3), 318-333.

Watanabe, Y. (2000). Washback effects of the English section of Japanese university entrance examinations on instruction in pre-college level EFL. *Language Testing Update, 27,* 42-47.

Watanabe, Y. (2004). Teacher factors mediating washback. In L. Cheng, Y. Watanabe, & A. Curtis

(Eds.), *Washback in language testing: Research contexts and methods* (pp. 129-146).

Mahwah, NJ: Lawrence Erlbaum.

Appendix A - Teacher questionnaire [English Version]

The purpose of this survey is to understand more about whether and how communicative language teaching (CLT) is perceived by Japanese teachers of English in high schools. Please read the sentences below and circle a number that best describes OC courses you teach. Choose only one response.

1. Check the academic year(s) when you teach/have taught OC course(s).
         [ ] 2010   [ ] 2009   [ ] 2008

2. Courses and grades you teach:
    Oral Communication 1   [1st grade   /   2nd grade   /   3rd grade   ]
    Oral Communication 2   [1st grade   /   2nd grade   /   3rd grade   ]

3. How many students are/were in the class?
                              (   ) students

4. Percentage of your students in the class are/were planning to take a university entrance exam that requires English subject?
        [ ] less than 10%   [ ]10 – 30 %   [ ]31 – 50 %
        [ ] 51 – 70 %   [ ]71 – 90 %   [ ] more than 91%

5. Percentage of your students in the class who will take the Center-Test
        [ ] less than 10%   [ ]10 – 30 %   [ ]31 – 50 %
        [ ] 51 – 70 %   [ ]71 – 90 %   [ ] more than 91%

6. Do you <u>usually</u> use English to conduct a lesson in the classroom?
      1. never   2. less than half   3. more than half   4. mainly in English

7. Do you agree that it is important to teach students to be able to speak English about daily topics?
      1. strongly disagree   2. disagree   3. agree   4. strongly agree

8. How often do you use the following teaching strategies/classroom activities? (Please rate your response on a 4 point scale where 1 = never; 2 = sometimes; 3 = often; 4 = always.)

	never	sometimes	often	always
(1) Chorus reading	1	2	3	4
(2) Multiple-choice grammar/vocabulary exercises	1	2	3	4
(3) Grammar-focused lecture	1	2	3	4
(4) Linguistic/Cultural lectures	1	2	3	4
(5) Pronunciation/accent reduction	1	2	3	4
(6) Listening exercises	1	2	3	4
(7) Pair/Group activities in English	1	2	3	4
(8) Individual presentations in English	1	2	3	4

                                        1   2

9. Do you think that the following factors have influence on the current course organization?
(Please rate your response on a 4 point scale where 1 = strongly disagree; 2 = disagree; 3 = agree; 4 = strongly agree.)

	strongly disagree	disagree	agree	strongly agree
(1) Curriculum (the Course of Study)	1	2	3	4
(2) University entrance examinations	1	2	3	4
(3) Term exams/Regional mock exams (*moshi*)	1	2	3	4
(4) Pre-teacher training experiences	1	2	3	4
(5) In-service teacher training experiences	1	2	3	4
(6) Students' insufficient speaking skills in English	1	2	3	4
(7) Teachers' insufficient speaking skills in English	1	2	3	4
(8) Number of students in class	1	2	3	4
(9) Textbooks				

10. Which of the following areas do you include in a term exam of OC courses you teach?
(Please rate your response on a 4 point scale where 1 = never; 2 = sometimes; 3 = often; 4 = always.)

	never	sometimes	often	always
(1)Reading comprehension	1	2	3	4
(2) Spelling/Vocabulary	1	2	3	4
(3) Grammar	1	2	3	4
(4) Translation	1	2	3	4
(5) Linguistic/Cultural knowledge	1	2	3	4
(6) Pronunciation/Stress	1	2	3	4
(7) Listening comprehension	1	2	3	4
(8) Speaking	1	2	3	4

11. How often do you have a speaking test <u>besides</u> a term exam?
        1. never   2. once in a year   3. 2~3 times in a year   4. more than 4 times in a year

12. What percentage of the final grade is allocated f the speaking test(s)?
                                [   ] percent

13. Please circle or check the item that best describes your background and current teaching situation.

1. Background information
     (1) Your gender: [ ] male   [ ] female
     (2) Your age: [ ] 22-29   [ ] 30-39   [ ] 40-49   [ ] above 50
     (3) Numbers of years teaching English in a high school:
        [ ] 0-3 years   [ ] 4-10 years   [ ] over 10 years
     (4) Your first language: [ ] Japanese   [ ] English   [ ] Other, specify: _____
     (5) Experience of living abroad in total:
        [ ] None   [ ] 1-6 months   [ ] 7-12 months
        [ ] 1-2 years   [ ] more than 3 years

2. Academic background
     (1) Your academic background:
        [ ] Bachelors   [ ] Masters [ ] PhD [ ] other, specify: _____
     (2) Your major
        [ ] English literature [ ] English linguistics [ ] Education
        [ ] TESOL [ ] Other, specify: _____
3. Experience of taking courses specifically in CLT
     (1) Pre-service [ ] yes   [ ] no
     (2) In-service   [ ] yes   [ ] no

- END OF QUESTIONNAIRE -
Thank you very much for your time and help.

Appendix B – Teacher Questionnaire [Japanese version]

<u>オーラルコミュニケーション(OC)コースについての質問</u>

ご担当になったＯＣのコースの状況についてお答え下さい。<u>複数のクラスをご担当の場合は、進度が早い方のクラスについてお答え下さい。</u>

１．ご担当の年度について、あてはまるもの全てにチェックをご記入ください。
      [   ]現在（2010 年度）   [   ] 2009 年度   [   ] 2008 年度

２．現在、または過去２年以内で一番最近に担当されたコースと学年：
      オーラルコミュニケーション１ [１年 /   ２年 /   ３年   ]
      オーラルコミュニケーション２ [１年 /   ２年 /   ３年   ]

３．ご担当ＯＣクラスの平均的な生徒数は何人ですか？
                                                           人

４．そのクラスで、英語の必要な大学受験を希望する（受験した）生徒はどのくらいの割合ですか？
   （先生のご経験からの推察で結構です。）
      [   ]10%以下   [   ]10 – 30 %   [   ]31 – 50 %
      [   ]51 – 70 %   [   ]71 – 90 %   [   ]91%以上

５．その中で、大学入試センターテストを受験する（した）生徒はどのくらいの割合ですか？（先生のご経験からの推察で結構です。）

      [   ]10%以下   [   ]10 – 30 %   [   ]31 – 50 %
      [   ]51 – 70 %   [   ]71 – 90 %   [   ]91%以上

６．ＯＣの授業（ALT 不在時）で、先生が英語を使われる割合はどの程度ですか？
      (1) ほぼ無い   (2) 半分以下   (3) 半分以上   (4) 主に英語 のみ

７．ＯＣの授業で、生徒が日常生活について英語で話せるようになるまで指導するべきだと思われますか？
      (1) 思わない   (2) あまり思わない   (3) やや思う   (4) 非常にそう思う

８．普段の OC の授業で以下のアクティビティを行う頻度はどのくらいですか？
   （１～４からひとつ選んでお答え下さい。）

	ほぼ無い	たまに行う	よく行う	毎回行う
(1) 声を合わせて音読	1	2	3	4
(2) 文法や語彙のクイズや小テスト	1	2	3	4
(3) 文法の解説	1	2	3	4
(4) 言語や文化の解説	1	2	3	4
(5) 発音やアクセントの訓練	1	2	3	4
(6) リスニング問題	1	2	3	4
(7) ペアやグループで行う英語のみでの活動	1	2	3	4
(8) 個人で行う英語でのプレゼンテーション	1	2	3	4

９．OCの授業を行う上で、以下の項目の影響を感じることはありますか？

	全くない	あまりない	ややある	非常にある
(1) カリキュラム及び学習指導要領	1	2	3	4
(2) 大学入試	1	2	3	4
(3) 定期考査や模試などのテスト	1	2	3	4
(4) 教員養成課程での訓練	1	2	3	4
(5) 現職研修で受けた訓練	1	2	3	4
(6) 生徒の英語能力の不足	1	2	3	4
(7) 教員の英語能力の不足	1	2	3	4
(8) クラスの人数	1	2	3	4
(9) ＯＣの教科書	1	2	3	4

１０．ＯＣの定期テストを作る際に、以下の問題はどの程度出題しますか？

	全く無い	ときどき	頻繁に	毎回必ず
(1) 英文読解	1	2	3	4
(2) スペリングや語彙	1	2	3	4
(3) 文法	1	2	3	4
(4) 英訳や日本語訳などの翻訳	1	2	3	4
(5) 文化や言語に関する知識	1	2	3	4
(6) 発音やアクセント	1	2	3	4
(7) リスニング	1	2	3	4
(8) スピーキング	1	2	3	4

１１．定期試験以外で年に何回ほどＯＣでスピーキング・テストを行いますか？
　(1) 全く無い　　　(2) 年に１〜２回　　　(3) 年に３〜４回　　　(4) 年に５回以上

１２．スピーキングテストの得点は、ＯＣの最終成績の何点程度ですか？
　　　　　最終成績 100 点中で[　　　　　] 点ほどをスピーキングテストに割り当てる。

１３．皆さんの条件に最も適しているものを選んでチェックマークをご記入ください。

1. バックグラウンドインフォメーション：
   (1) 性別：[   ] 男性　　　[   ] 女性

   (2) 年齢：[   ] 22-29　　[   ] 30-39　　[   ] 40-49
   　　　　　　[   ] 50 歳以上

   (3) 高校での英語教員としてのご勤務経験年数：
   　　　　　[   ] 0-3 年以内　　　[   ] 3-10 年▪内　　　[   ] 10 年以上
   　　　　　[   ] 常勤講師：_____年目　　[   ] 非常勤講師：_____年目

   (4) 第一言語：[   ] 日本語　[   ] 英語　[   ] その他：＿＿＿＿＿＿＿＿

   (5) 英語圏に滞在された年数：
   　　　　　[   ] 0-1 年以内　　　　　　[   ] 1〜3 年以内　　　[   ] 3〜5 年以内
   　　　　　[   ] 5 年以上

2. 学歴及び専攻に関する情報：
   (1)学位または職業資格について（J-shine 資格などあればお答えください）
   　[   ] 学士　　[   ] 修士　　[   ] 博士　　[   ] その他：＿＿＿＿＿

   (2)大学での専攻について
   [   ] 英文学　[   ] 英語学　[   ] 教育学　[   ] TESOL
   [   ] その他：＿＿＿＿＿＿＿＿

3. Communicative Language Teaching(CLT)の経験に関する情報：
   (1) CLT に特化したコースを学生時代に受講されたことがありますか？
   　　　　　　[   ] はい　　　　　　[   ] いいえ

   (2)CLT に特化したコースを卒業後に受講されたことがありますか？
   　　　　　　[   ] はい　　　　　　[   ] いいえ

―アンケートはここで終了になります―

お忙しい中、ご協力を本当にありがとうございました

Appendix C- Classroom Observation Guide

School:	Date:	Observer: Rika Tsushima
Teacher:	Grade level:	Number of students:
Subject: Oral Communication I / Ⅱ	Class length:	

Observation Form
Focus on Speaking Events

Description of Classroom

Event No.	Time	Description of task types	Focus of tasks		Participants organization			Language			Materials			Teacher talk and behavior	Student talk and behavior
			form	meaning	Class	Group/Pair	Individual	English	Japanese	mixed	Text	Audio	Visual		

Adapted from Tuner's (2000) classroom observation guide

Appendix D - Terms Exams Analysis Guide for OC courses

	Focus of the question									score	language of the questions			language of the answers			Description
	pronunciation/accent	spelling/ vocabulary	reading comprehension	listening comprehension	speaking	writing	translation	grammatical knowledge	sociolinguistic knowledge		Japanese	English	Mixed	Japanese	English	Mixed	
Q1																	
Q2																	
Q3																	
Q4																	
Q5																	
Q6																	
Q7																	
Q8																	
Q9																	
Q10																	
total										100/ 100							

Appendix E – Teacher Interview Questions [English Version]

I. Questions about Oral Communication courses

Q1. What do you think is important for students to learn in the courses?

Q2. How often does an assistant language teacher (ALT) teach in your class?

Q3. When you have an ALT, do you change the classroom organization from usual ones? If so, why?

Q4. How do you differentiate activities from one you do in English to one you do not?

Q5. What are the difficulties you have to conduct OC lessons?

II. Questions about Oral Communication courses

Q6. How often do you test students' speaking skills?

Q7. Describe your testing and evaluation strategies in OC courses.

Q8. Who usually assesses your students' speaking skills? (Do you cooperate with the ALT to test/evaluate students?)

Q9. When you test/evaluate students' speaking skills, what do you keep in your mind?

Q10. Do you think your test is congruent with the goal of the curriculum?

III. Questions about inclusion of a speaking component in entrance exams

Q11. Have you had students who needed preparation for a speaking test to pass an entrance exam? If so, what did you do to train the student?

Q12. If the Center Test included a speaking component, what consequence on your teaching do you think you would have at first?

Q13. If the Center Test includes it, would you change your testing and evaluation strategies? If so, what would you do?

Appendix F – Teacher Interview Questions [Japanese Version]

I. OC の授業に関する質問

１．ＯＣのコースで生徒が学ぶべきことは何だと思われますか？

（スピーキング力について）どの程度まで、高校生は到達するべきだと思いますか？（挨拶程度、

旅行ができる、日常会話、知的な会話ができる）

２．ＡＬＴが授業に来る機会はどの程度ですか？その場合の授業の構成はいつもと違うものになります

か？その理由は？

３．ＯＣの授業で、英語を使用する割合はどの程度ですか？英語を使用する活動と、日本語を使用する

活動、どうやって言語を使い分けていますか？

４．先生が理想とするＯＣの授業とはどのようなものですか？その理想とするモデルと、現実は一致し

ているでしょうか？

５．もし、一致していない場合は、障害となっているものは何でしょうか？思いつくまま、挙げてくだ

さい。

II. テストに関する質問

１．生徒のスピーキング力を測るテストを行うことがありますか？

２．スピーキングテストを行う場合、その評価を行うのは誰ですか？

３．テストの方法と、その評価方法を詳しく教えてください。

４．スピーキングの評価基準を決める場合、何に主に注意して評価しますか？

５．ご自分の作るテストは、ＯＣコースの指導目標と一致していると思われますか？

III. 入試制度に関する質問

１．スピーキングの必要な試験を受けた生徒を担当したことがありますか？

２．今後、ＡＯ入試を初め、スピーキングが必要な大学入試が増える傾向にありますが、それについて

どう思われますか？

３．もし、センターテストがスピーキング問題を含めた場合、どのような影響が考えられますか？