

**EFL EXAMINATION WASHBACK IN JAPAN: INVESTIGATING
THE EFFECTS OF ORAL ASSESSMENT ON TEACHING AND
LEARNING**

**A dissertation submitted in part fulfilment of the requirements for the degree of
Master in Education in the Faculty of Education**

2005

NICHOLAS A. CAINE

**Language and Literacy Studies in Education
The University of Manchester**

ABSTRACT

The influence a test has on teaching and learning is commonly referred to as the ‘washback effect’. It is often argued that English examinations in Japan, which tend to be heavily grammar-orientated, have a negative washback effect on teaching and learning. The areas of writing and speaking are of particular concern as both tend to be assessed via indirect testing methods. This study, then, examines the effects of existing English tests in Japan and also proposes an original direct test of speaking, which is subsequently trialled in a sample learning context. An attempt is then made to determine the extent and nature of washback resulting from this new speaking test.

Chapter 1 looks at the study of English as a foreign language (EFL) within the context of Japanese high schools. In particular, the discussion focuses on the ‘mismatch’ that occurs between the levels of curriculum planning and actual classroom implementation. It is suggested that one of the reasons behind the apparent failure of the official communicative syllabus adopted by the Japanese Ministry of Education (MEXT) is due to the existence of the “hidden syllabus” driven by the content of EFL examinations.

Chapter 2 examines the phenomenon of washback with a review of the published literature in the field. The notions of test washback and test impact are introduced before the discussion goes on to look at the different kinds of influence a test might have, with specific reference to the ‘Washback Hypothesis’ put forward by Alderson & Wall (1993). The issue of washback occurring as a result of ‘high stakes’ testing is also considered and finally, the question of whether positive washback can be nurtured and thus improve curricula is analysed in greater detail.

Chapter 3 presents the rationale for the design of the new speaking test. Firstly, however, a brief background to communicative language testing is presented in order to consider

an approach that conforms to the philosophy of the official English language syllabus in Japan. An outline of current EFL tests faced by the learners in the sample is also offered in this chapter.

Chapter 4 details the data collection techniques being used to measure the washback effect of existing tests and the new speaking test. An outline of the research context is also provided along with details of the learners and teachers taking part in all aspects of the study.

Finally, Chapter 5 discusses the research findings and looks at the implications for both EFL test design, and for the future implementation of a communicative syllabus in the Japanese context.

CONTENTS

CHAPTER 1: ENGLISH AS A FOREIGN LANGUAGE IN JAPAN	9
1.1 Official language policy: the syllabus goals of the Ministry of Education	9
1.2 Dominant methodology	10
1.3 The hidden syllabus: factors affecting the implementation of communicative language teaching	11
CHAPTER 2: LITERATURE REVIEW	15
2.1 Interpretations of test washback and test impact	15
2.1.1 What is test washback?	15
2.1.2 What is test impact?	16
2.2 The influence of tests on teaching and learning	19
2.2.1 The Washback Hypothesis	19
2.2.2 Washback as a result of ‘high stakes’ testing	22
2.2.3 Positive washback: can it be nurtured and can it improve curricula?	24
CHAPTER 3: APPROACHES TO LANGUAGE TESTING	29
3.1 Communicative language testing	29
3.2 Current tests of English in the teaching and learning context	32
3.3 Rationale for the new speaking test	39
CHAPTER 4: TEST WASHBACK WITHIN THE SAMPLE TEACHING AND LEARNING CONTEXT	45
4.1 The teaching and learning context	46
4.1.1 Teachers in the sample	47
4.1.2 Learners in the sample	48
4.2 Measuring the washback effect of current EFL tests in the sample context	48
4.2.1 Classroom observation	48
4.2.2 The survey instrument: teacher and student questionnaires	52
4.3 Trialling the speaking test and measuring washback	54
CHAPTER 5: RESEARCH FINDINGS AND IMPLICATIONS	56
5.1 Classroom observation results and discussion	56
5.2 Survey results and discussion	63
5.2.1 Teacher questionnaire	63
5.2.2 Student questionnaire #1	72
5.2.3 Student questionnaire #2	78
5.3 Conclusion and implications	81

APPENDIX I: National Center for University Entrance Examinations (NCUEE) ‘Center Test’ 2004 sample questions	85
APPENDIX II: The Society for Testing English Proficiency (STEP) performance descriptors for grades 1 – 5	97
APPENDIX III: New speaking test and sample materials	99
APPENDIX IV: Analytic marking scheme for the speaking test	106
APPENDIX V: Communicative Orientation of Language Teaching (COLT) Observation Scheme coding sheet	107
APPENDIX VI: Results from the classroom observation sessions using the COLT (Part A) Observation Scheme	108
APPENDIX VII: Teacher questionnaire	109
APPENDIX VIII: Student questionnaire #1	116
APPENDIX IX: Student questionnaire #2	120
BIBLIOGRAPHY	122

LIST OF TABLES AND FIGURES

TABLES

CHAPTER 3

TABLE 3.1: The 5-point grading scale	44
---	-----------

CHAPTER 4

TABLE 4.1: Outline of the groups involved in the observation sessions	49
--	-----------

TABLE 4.2: Outline of the groups involved in the questionnaire survey	53
--	-----------

CHAPTER 5

TABLE 5.1 Patterns of participant organisation adopted by teacher A (% of total class time)	58
--	-----------

TABLE 5.2 Patterns of participant organisation adopted by teacher B (% of total class time)	59
--	-----------

TABLE 5.3 Student modality during teacher A's classes (% of total class time)	61
--	-----------

TABLE 5.4 Student modality during teacher B's classes (% of total class time)	61
--	-----------

TABLE 5.5 Teacher questionnaire: Results from Likert scale questions 4 – 11	65
--	-----------

TABLE 5.6 Teacher questionnaire: Results from Likert scale questions 12 – 14	66
---	-----------

TABLE 5.7 Student questionnaire #1: Results from Likert scale questions for group 4 (n=26)	73
---	-----------

TABLE 5.8 Student questionnaire #1: Results from Likert scale questions for group 5 (n=20)	74
---	-----------

TABLE 5.9 Results of student questionnaire #2 (Group 4, n=27)	79
--	-----------

FIGURES

CHAPTER 2

FIGURE 2.1 Stakeholders in the testing community	16
---	-----------

CHAPTER 3

FIGURE 3.1 The pattern of interaction for the proposed speaking test	41
---	-----------

CHAPTER 5

FIGURE 5.1 Patterns of participant organisation during the four observation sessions	59
---	-----------

FIGURE 5.2 Comparison of student modality during the four observation sessions	61
---	-----------

(1) Copyright in the text of this dissertation rests with the Author. Copies (by any process) either in full, or of extracts, may be made **only** in accordance with instructions given by the Author. This page must form part of any such copies made. Further copies (by any process) of copies made in accordance with such instructions may not be made without the permission (in writing) of the Author.

(2) The ownership of any intellectual property rights which may be described in this dissertation is vested in the University of Manchester, subject to any prior agreement to the contrary, and may not be made available for use by third parties without the written permission of the University, which will prescribe the terms and conditions of any such agreement.

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my supervisor, Richard West, for his guidance and constructive feedback throughout all the various stages of this dissertation. I would also like to thank my family for their support, particularly my father, Michael, for devoting so much of his time to proof-reading each chapter. Special thanks must also go to my wife, Natsu, not only for her patience and encouragement during the many hours I spent working at the computer, but also for her assistance with Japanese translations. Finally, I am most grateful to my English teaching colleagues and the students at Keitoku High School, Imari, for their participation in the project. Particular thanks here must go to Hidetsugu Shima and Eishi Sadakata for kindly allowing me to observe their classes and, of course, to those students who took part in the paired speaking tests.

Chapter 1: English as a Foreign Language in Japan

1.1 Official language policy: the syllabus goals of the Ministry of Education

The Japanese Ministry of Education – which forms part of the all-encompassing Ministry of Education, Culture, Sports, Science and Technology (MEXT¹) – has, for some time now, promoted the use of communicative teaching methodology in the English classroom. In support of this philosophy, the syllabus guidelines document produced by MEXT (The Course of Study for Foreign Languages 1993) stresses the importance of developing students' ability to use English for the purpose of everyday communication.

An Action Plan to Cultivate “Japanese with English Abilities” (MEXT 2003) also highlights the measures that need to be taken to improve second language education in Japan. This document presents a series of goals, which, it is hoped, can be achieved within a five-year period. The following statement is made in relation to the required English abilities of students graduating from high school:

On graduation from a senior high school, students can conduct normal communication with regard to topics, for example, relating to daily life. (English language abilities for graduates should be the second level or the pre-second level of the Society for Testing English Proficiency (STEP²) on average).

(MEXT 2003: 1)

With regard to actual teaching methodology, the paper reiterates the Ministry's policy of promoting communicative language teaching (CLT) by going on to state that ‘instruction mainly based on grammar and translation or teacher-centered classes is not recommended’ (MEXT 2003: 3).

¹ MEXT (The Ministry of Education, Culture, Sports, Science and Technology (Monbu-kagaku-sho)) came into being in January 2001. Prior to that, the Ministry of Education was known as Monbusho (sometimes referred to as Mombusho).

² Established under the guidance of the Japanese Ministry of Education in 1963, the STEP (or Eiken) is an optional seven level proficiency test, administered 3 times a year exclusively to Japanese learners of English. A more in-depth analysis can be found in Chapter 3 (pp. 35 – 36).

Whilst the aims and objectives of MEXT represent worthy statements of intent, the degree to which they are actually being realised is somewhat open to question. Indeed, it could be some time before the majority of high school graduates are able to reach the level of proficiency demanded in the Ministry's Action Plan. Despite the fact that English is studied through six years of secondary education – starting from the age of twelve and continuing up until graduation from high school at eighteen – many learners continue to experience difficulty in using the language for purposeful communication, often encountering problems in the areas of speaking and writing. Thus, it would appear that the primary objectives of the syllabus prescribed by MEXT, along with the goals put forward in the Action Plan remain, as yet, largely unfulfilled.

One of the main reasons behind the apparent failure of both the syllabus and the Action Plan perhaps lies in the demands placed upon teachers and learners by the 'hidden syllabus'. Before discussing this in greater detail, however, let us first of all consider the kind of classroom environment that has come to typify the Japanese approach to English language education over the years.

1.2 Dominant methodology

The study of English as a foreign language (EFL) is often associated with a very 'traditional' pedagogy in Japan and has become synonymous with:

- a teacher-centred approach that relies heavily on grammar-translation and rote learning methods;
- learning materials which emphasise language usage as opposed to actual language use, resulting in lessons that, for the most part, focus on discrete-point grammar items and the translation of reading passages into the first language;
- large classes, often numbering 40 or more students of mixed ability;
- classrooms which, in terms of layout, arguably serve to promote the preferred

‘lockstep’ approach, with students usually seated in rows consisting of single desks positioned one behind the other.

This traditional approach is sometimes referred to as yakudoku (Gorsuch 1998), which roughly translates as “the grammar-translation reading method” (Brown 2002) and, with many classrooms continuing to exhibit such characteristics, EFL in Japan is often criticised on the grounds that it has been slow to progress. Indeed, as the calls for a more communicative approach increase, there has been much debate with regards to the best way to achieve this objective and thus improve the communicative competence of learners, which represents the ultimate target of the Ministry of Education.

1.3 The hidden syllabus: factors affecting the implementation of communicative language teaching

There is still much disagreement – and to some extent, confusion – over the most effective way to approach the study of English in Japan, especially in terms of content (i.e. what should be taught) and methodology (i.e. how it should be taught). On the one hand, there are those who believe that the more traditional ‘lockstep’ approach, as outlined in section 1.2, is most effective for meeting the needs of Japanese learners and the system as a whole. On the other hand, it is felt that communicative language teaching offers the best way forward if English education in Japan is to demonstrate any real signs of progress. Watanabe (2004: 132), for example, notes that:

...now, in response to the need for international communication, there is a greater emphasis on active language use involving the exchange of both spoken and written information.

This would certainly appear to be the case at the level of curriculum planning where the Ministry of Education is keen to promote CLT. In practice, however, greater importance continues to be placed upon learning about the language rather than on actual language

use – a situation ideally suited to the more traditional grammar-translation approach still preferred by so many Japanese teachers of English.

It becomes increasingly clear, then, that the communicative approach faces considerable opposition and that several barriers have yet to be overcome if CLT is to be fully integrated into the English syllabus. Not surprisingly, much of the opposition comes from teachers who are more familiar with grammar-translation methodology having received little, if any, training in communicative techniques – a point confirmed by Reesor (2003: 63):

Most JTEs (Japanese Teachers of English) learned English through the grammar-translation method and as a result, this is the methodology which they feel most comfortable with as teachers.

If no training in communicative teaching methodology is provided during initial training courses, it creates something of a dilemma for many Japanese teachers of English, who must then decide whether to gain knowledge in the area through additional, independent study. Sakui (2004) also makes the observation that there appears to be no agreed definition of communicative language teaching at either an instructional level or a conceptual level, thus adding to the problems of implementation.

Teacher perceptions of communicative language teaching are in many ways influenced by the existence of a ‘hidden syllabus’. Within the Japanese high school context, this hidden syllabus is undoubtedly shaped by the content of English examinations – particularly those administered by universities as part of their entrance procedures. Johnson (1989: 6) argues that, if a syllabus is to have credibility, then subsequent test items must be related to the content of that syllabus, and that ‘item types in examinations need to be selected and constructed with this ‘washback’ effect in mind’. Unfortunately, university entrance examinations in Japan are notoriously grammar-

orientated; somewhat contradicting the syllabus aims of the Ministry of Education. Sakui (2004: 159), for example, suggests that:

...contrary to the teachers' aspirations to incorporate CLT into their teaching, they cannot ignore the demand to prepare students for entrance examinations.

It is important for test designers then, to be aware of the possibility of washback (positive or negative) – defined by Hughes (1989: 1) as ‘the effect of testing on teaching and learning’. Brown (2002) makes the observation that if official syllabus policy in Japan continued to recognise grammar-translation as the preferred method of instruction – as indeed used to be the case – then university examinations would have greater validity (i.e. test content would be more consistent with syllabus content). However, following the introduction of the new syllabus guidelines and the promotion of communicative methodology, ‘a mismatch was created between the yakudoku entrance examinations and any curricula that had actually responded to Mombusho’s guidelines’ (Brown 2002). This would seem to suggest that in the current climate, entrance examinations are exerting unfavourable washback upon the communicative syllabus and the degree to which it can be successfully implemented. This, in turn, is also arguably having a detrimental effect upon teaching (and learning) practices.

The aim of this study is to test this hypothesis and establish whether examination washback is indeed occurring in the context outlined throughout this introductory chapter. In particular, the following research questions will be considered:

1. To what extent does test content influence teaching methodology?
2. Can teaching and learning be improved by adopting direct testing techniques, thus eliminating the potential for mismatch that arises from the communicative syllabus aims of MEXT?

The discussion proceeds with a review of the literature relevant to the phenomenon of washback (Chapter 2). Chapter 3 will explore in greater detail some of the problematic

aspects of existing tests and offer an alternative test design. The new test proposes to measure speaking skills directly – at present, the productive skills of speaking and writing, if tested at all, tend to be assessed by indirect methods. Chapter 4 will look at the trial of the test and will also provide details of the primary methods of data collection being used to address the two research questions – namely questionnaires and classroom observations. Finally, Chapter 5 offers a discussion of the results in relation to the research questions and considers the implications for CLT implementation and EFL test design in the Japanese context.

Chapter 2: Literature Review

2.1 Interpretations of test washback and test impact

2.1.1 What is test washback?

Numerous explanations of the term ‘washback’ can be found throughout the published research and literature on language testing. One of the most common definitions sees the concept referred to as the influence of testing on teaching and learning (e.g. Alderson & Wall 1993; Gates 1995; Cheng & Curtis 2004). Similarly, Shohamy et al (1996: 298) define washback as ‘the connections between testing and learning’ and Saville (2000: 4) and Hughes¹ (1989: 1) as ‘the effect of testing on teaching and learning’. Messick (1996: 241), noting that washback can have either harmful or positive effects, describes 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’.

Such definitions refer only to washback in terms of the influence that tests might have upon the language classroom and the participant roles of teacher and learner. In other analyses a much broader interpretation is offered, taking the view that tests can have more far-reaching effects within the field of education. For example, Andrews (2004: 37), in an article that explores the relationship between washback and curricular innovation, looks beyond the classroom, and uses the term to describe ‘the effects of tests on teaching and learning, the educational system, and the various stakeholders in the education process’. The view of washback presented here approaches what some writers differentiate as test impact.

¹ Whilst ‘washback’ is the preferred term in British applied linguistics (Alderson & Wall, 1993: 115), some writers such as Hughes refer to the same phenomenon as ‘backwash’.

2.1.2 What is test impact?

Although the terms washback and impact are sometimes used synonymously – as indeed demonstrated by Andrew’s (2004) definition of the former, above – test impact more accurately refers to the wider implications and effects of a given test. For example, whilst acknowledging that washback can be seen as the influence and effect of tests on teaching and learning, McNamara (2000: 74) notes that:

Tests can also have effects beyond the classroom.
The wider effect of tests on the community as a whole,
including the school, is referred to as test **impact**.

Wall (1997: 291) makes a similar distinction between test washback and test impact, agreeing that the term ‘impact’ more accurately refers to:

...any of the effects that a test may have on
individuals, policies or practices, within the classroom,
the school, the educational system or society as a whole.

Taylor (2000: 2), building upon a model proposed by Rea-Dickins (1997), who ‘identified at least 5 stakeholder categories: learners, teachers, parents, government and official bodies, and the marketplace’, offers a more detailed conceptualisation in order to illustrate the wider societal effects of a test (i.e. test impact):

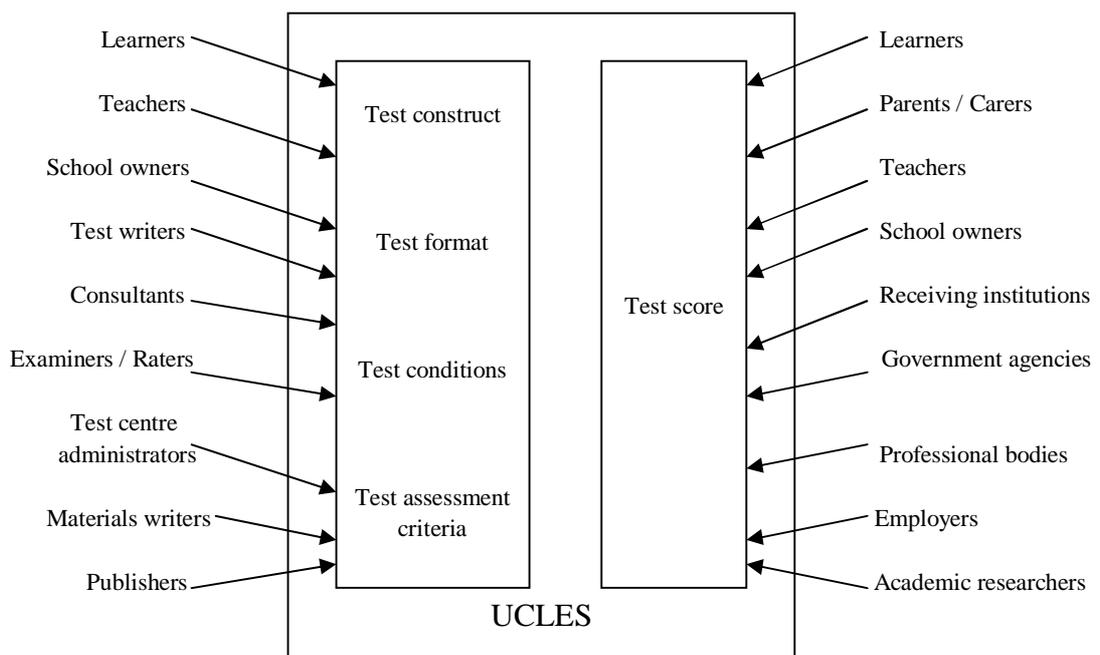


Figure 2.1 Stakeholders in the testing community

Note: UCLES = University of Cambridge Local Examinations Syndicate (see following discussion)

The above model provides a useful illustration of the fact that a test can have impact upon the various stakeholders involved, at different points in the testing process:

Some of the stakeholders listed above (e.g. examiners and materials writers) are likely to have more interest in the ‘front end’ of a test, i.e. the test assessment criteria or test format. Others may see their stake as being primarily concerned with the test score. Some stakeholders, such as learners and teachers, will naturally have an interest in all aspects of the test.

(Taylor 2000: 2)

Although referring specifically to work carried out by the University of Cambridge Local Examinations Syndicate (UCLES) on the Cambridge suite of EFL examinations (hence the inclusion of that particular examination body in figure 2.1), the comprehensive model presented by Taylor (2002: 2), depicting the various stakeholders on whom a test has impact – what Saville (2000: 4) calls the ‘language testing constituency’ – is of equal relevance to the Japanese context under discussion in this study. Thus, examination bodies such as the Society for Testing English Proficiency (STEP) who administer the national Eiken proficiency test; the National Center for University Entrance Examinations (NCUEE), who are responsible for the Center Test used in the preliminary stages of university entrance procedures; or the American-based Educational Testing Service (ETS), who administer the Test of English for International Communication (TOEIC) and the Test of English as a Foreign Language (TOEFL) tests; could be applied to the above model.

Bailey (1996: 263-264) also adopts a more holistic view regarding the effects of tests, but prefers to consider overall impact in terms of ‘washback to the learners’ and ‘washback to the programme’. In the latter group she includes ‘teachers, administrators, curriculum developers, counsellors, etc.’ (Bailey 1996: 264) – groups of individuals who, according to Taylor (2000), would be more likely to have an interest in the ‘front end’ of

a test (see figure 2.1). In view of the research questions outlined at the end of chapter 1 which relate firstly, to the influence that existing tests have upon teaching (programme washback), and secondly, to the possibility of a new test design bringing about improvement to teaching and learning whilst eliminating the potential for syllabus mismatch (programme and learner washback), the model of washback presented by Bailey (1996) offers a useful framework for the purposes of this study.

Throughout this initial discussion it becomes clear that the term ‘washback’ is open to a variety of interpretations and that there are a number of important variables to consider when conducting research into the issue. For example, how are we defining the term ‘washback’ exactly? Are we using the term to describe the effects of a test on teaching and learning only, or are we using it in the wider sense to include the effects of a test on other stakeholders in the education process (as illustrated in figure 2.1), which as noted above would be more accurately referred to as test impact? Other issues (which will be addressed in the following sections) also arise concerning the kind of influence that a test might have. In addition, if – as suggested by Messick (1996: 241) – the effects of a test can be harmful or beneficial, is it possible to foster positive washback, and how might we achieve this objective?

It is not within the scope of this study to look in detail at the wider implications of testing – as already mentioned we will be primarily concerned with the two areas identified by Bailey (1996: 263-264) – i.e. ‘washback to the programme’ and ‘washback to the learners’. In other words, I will adopt the narrower definition of washback by concentrating on the effects that a test has on teaching and learning.

Let us now look at the Washback Hypothesis (Alderson & Wall 1993) in order to consider how a test might affect – or influence – the language teaching and learning environment.

2.2 The influence of tests on teaching and learning

2.2.1 The Washback Hypothesis

The title of an article written by Alderson & Wall (1993) posed the following question: “Does Washback Exist?” In the article, it was noted that whilst many assertions had been made relating to the influence of tests, there was very little in the way of empirical evidence to back up the claims that tests affected teaching and learning, in either a negative or positive way. In order to examine in greater depth the relationship between washback and ‘influence’, Alderson & Wall (1993: 120-1) put forward the 15 hypotheses listed below, highlighting more specifically some of the ways in which a test might affect teaching and learning. Bailey (1996: 265-266) notes that five of the hypotheses relate to ‘washback to the learners’ and six relate to ‘washback to the programme’. Thus, the letters in parentheses have been added in order to illustrate this point, with ‘L’ indicating those statements concerning washback to the learners, ‘P’ to those concerning washback to the programme.

Some possible Washback Hypotheses

- (1) A test will influence teaching. **(P)**
- (2) A test will influence learning. **(L)**
- (3) A test will influence **what** teachers teach **(P)**; and
- (4) A test will influence **how** teachers teach **(P)**; and therefore by extension from (2) above:
- (5) A test will influence **what** learners learn **(L)**; and
- (6) A test will influence **how** learners learn. **(L)**

- (7) A test will influence the **rate** and **sequence** of teaching (**P**); and
- (8) A test will influence the **rate** and **sequence** of learning. (**L**)
- (9) A test will influence the **degree** and **depth** of teaching (**P**); and
- (10) A test will influence the **degree** and **depth** of learning. (**L**)
- (11) A test will influence attitudes to the content, method, etc. of teaching and learning (**P**)
- (12) Tests that have important consequences will have washback; and conversely
- (13) Tests that do not have important consequences will have no washback.
- (14) Tests will have washback on **all** learners and teachers.
- (15) Tests will have washback effects for **some** learners and **some** teachers, but **not** for others.

Noting the uncertain nature of the phenomenon, however, Alderson & Wall (1993: 117) observe that:

The Washback Hypothesis **seems to assume** that teachers and learners do things they would not necessarily otherwise do because of the test.

(emphasis added)

They also make the point that even a ‘poor’ test could have a positive washback effect if it encouraged motivation on the part of learners or teachers. For example, a test might encourage learners to ‘do their homework, take the subject being tested more seriously, and so on’, whereas teachers might ‘prepare lessons more thoroughly’ (Alderson & Wall 1993: 117). It is also noted that a ‘good’ test, on the other hand, could have adverse effects by bringing about learner anxiety or, in the case of teachers:

...the fear of poor results, and the associated guilt, shame, or embarrassment, might lead to the desire for their pupils to achieve high scores in whatever way seems possible.

(Alderson & Wall 1993: 118)

If teachers are unduly influenced by a test this could have serious implications if the content of that test fails to reflect the aims and objectives of the prescribed syllabus – the

result being that significant areas of the syllabus which are not tested may then be seen as irrelevant and, as a consequence, will be neglected. This ‘narrowing of the curriculum’ (Madaus 1988, cited in Wall 2000: 500) is clearly related to the ‘hidden syllabus’ discussed in chapter 1, which tends to be driven by the content of university entrance examinations in Japan.

Despite the lack of empirical research into the phenomenon of washback, Alderson & Wall are able to cite some studies that had been conducted prior to the publication of their article in 1993. Of particular note are those conducted by Kellaghan et al. (1982), Wesdorp (1982) and Hughes (1988) – although it should be pointed out that the former was a general education study and not specific to language education. In their ensuing discussion, it is clear that evidence of either beneficial or harmful washback (and the extent to which it had occurred within a particular context) was often tenuous – remaining unproven or, at best, inconclusive. For example, to take the Kellaghan et al. (1982) study (which looked at the impact of introducing standardised tests in Irish schools) as a case in point, Alderson & Wall (1993: 122) make two critical observations. Firstly they note that, whilst the results were ‘beneficial to the schools, or at least not harmful’, the conditions for the experiment were artificial, particularly in view of the fact that the test would not be used to determine students’ future educational or employment prospects. Thus, it was less likely that the study would uncover any evidence of negative washback. Secondly, the failure to use direct observation as a method of data collection was regarded as another drawback – the result being that that the research relied too heavily upon anecdote, assertion, or interviews and surveys.

Indeed, Alderson & Wall (1993: 123) regard the absence of classroom observation as a common problem with much of the research that had been conducted into washback up until that point in time. The implication is that observation should be used not in

isolation, but rather as part of a combination of complementary research techniques. Indeed, it is notable that subsequent studies in the area of washback in language testing have tried to address this issue by employing a more varied use of data collection procedures: examples include Watanabe (1996) and Cheng (1999) who both used classroom observation in conjunction with teacher interviews; Cheng (1997), who used a combination of classroom observation, teacher questionnaires and student questionnaires, and Burrows (2004) who conducted teacher surveys and interviews, in addition to carrying out classroom observations.

2.2.2 Washback as a result of ‘high stakes’ testing

According to the ‘Washback Hypothesis’ (Alderson & Wall 1993: 120-1) ‘tests that have important consequences will have washback’ (see page 20 of this discussion). In other words, tests will arguably have a greater influence on teaching and learning in a ‘high-stakes’ situation – i.e. one in which the test is typically used ‘to compare and rank individuals, schools or national systems’ (Chapman & Snyder Jr. 2000: 458) and whose ‘primary use is to ration future opportunity as the basis for determining admission to the next layer of education or to employment opportunities’ (ibid.). As Andrews (2004: 37) also observes:

It is precisely the power of high-stakes tests (or the strength of the perceptions which are held about them) that makes them potentially so influential upon the curriculum and curricular innovation.

In chapter one, it was suggested that the failure of Japanese high schools to implement fully the communicative syllabus recommendations of MEXT was due to the powerful washback effect of ‘high stakes’ EFL tests. In particular, it was felt that the demands placed upon teachers and learners by the university entrance examinations were a major cause of this curriculum mismatch.

The research findings of studies that have involved so-called ‘high-stakes’ tests confirm some evidence of washback, although it is often reported that, in the case of teaching, such tests have a greater effect upon content and rather less upon the actual methodology employed by teachers. For Example, Cheng (1997), who looked at the effect that changes to the HKCEE² in English had upon teaching and learning, found that washback occurred, but only in terms of bringing about change to teaching materials. Evidence that the changes to the test brought about changes to the way teachers taught was not conclusive.

Likewise, in the context of Japan, research conducted by Watanabe (1996; 2004) examining the belief that teachers’ reliance on grammar-translation comes about as a direct result of university examination content, shows that the examinations exert a washback effect on some teachers but not on others. His findings suggest that factors such as the educational background, personal beliefs and experience of the individual teacher have as much influence over which methodology is employed in the classroom (Watanabe 1996: 330-1).

Chapman & Snyder Jr. (2000: 462) also question the extent to which high-stakes testing influences teachers’ classroom methodology. Citing a general education study from Uganda by Snyder et al. (1997), which found that changes made to a national examination did not have the desired effect of encouraging teachers to alter their instructional practices, they suggest that ‘it is not the examination itself that influences teachers’ behavior, but teachers’ beliefs about those changes’ (Chapman & Snyder Jr. 2000: 462).

² The Hong Kong Certificate of Education Examination

2.2.3 Positive washback: can it be nurtured and can it improve curricula?

Hughes (1989: 44-46) suggests seven ways in which positive washback might be achieved:

1. Test the abilities whose development you want to encourage.

Hughes (1989: 44) notes that ‘there is a tendency to test what it is easiest to test rather than what it is most important to test’. Thus, there are obvious implications here for the assessment of productive language skills (i.e. writing and speaking) when subjective marking is required. A criticism often heard in Japan is that examinees are often required to produce very little – if any – actual language in order to pass an EFL examination such as the STEP, TOEIC or the Center Test, with the emphasis very much placed on speed (in terms of reporting scores to candidates’ selected universities before the second stage of the entrance procedure) and reliability.

2. Sample widely and unpredictably.

To avoid the content of a test becoming too predictable, item types should be designed so that a range of tasks are utilised. A major feature/criticism of the university Center Test (see Appendix I) is that it tends to follow the same format every year.

3. Use direct testing.

Hughes (1989: 45) defines direct testing as ‘the testing of performance skills, with texts and tasks as authentic as possible’ and says that ‘immediately we begin to test indirectly, we are removing an incentive for students to practise in the way that we want them to’ (ibid.). As already mentioned, productive English skills are rarely assessed in the Japanese context.

4. Make testing criterion-referenced.

Criterion-referenced tests are seen to exhibit beneficial washback due to their motivating factors. For example, rather than being measured against the performance of other test-takers, Hughes (1989) believes that it is better when candidates are clear about exactly what they have to achieve in order to pass a test.

5. Base achievement tests on objectives.

Here, Hughes (1989: 46) argues that achievement tests should be based on objectives, as opposed to 'detailed teaching and textbook content' in order that teaching and learning can be evaluated against those objectives.

6. Ensure (that the) test is known and understood by students and teachers.

In order for beneficial washback to be fully realised, 'the rationale for (a) test, its specifications, and sample items should be made available to everyone concerned with preparation for the test' (Hughes 1989: 46). Hughes (ibid.) notes that this takes on particular importance during the introduction of a new test that perhaps contains new testing methods.

7. Where necessary, provide assistance to teachers.

When a new test is introduced in order to foster positive washback, perhaps with the intention of ultimately changing a teacher's preferred methodology in some way, then it is important that teachers affected by that test are fully conversant with the new approach being adopted. Hughes (1989: 46) uses the example of a communicative language test replacing a more traditional grammatical and vocabulary format and notes that:

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. It may simply cause chaos and disaffection.

It has been suggested elsewhere in this paper that, generally speaking, non-native teachers of English in Japan receive little training in communicative methodology. Therefore, teacher-training programmes must also be a consideration if we are planning to introduce a communicative language test.

Hughes concludes the discussion by admitting that many of his suggestions could be expensive to implement and, as a consequence, be seen as impractical. It is noted that

practicality is also included among ‘the desirable qualities’ (Hughes 1989: 47) of a test, thus, ‘it is good that a test should be easy and cheap to construct, administer, score and interpret’ (ibid.). Nevertheless, Hughes (1989: 47) argues that the pursuit of positive washback should remain a primary objective in language test design:

Before we decide that we cannot afford to test in a way that will promote beneficial backwash, we have to ask ourselves a question: what will be the cost of **not** achieving beneficial backwash?

To reiterate, EFL tests in Japan, particularly the university Center Test, whilst addressing reliability and practicality issues so thoroughly, tend to do so at the expense of test validity. Commenting on this situation, Brown (1995: 277) notes that entrance examinations – i.e. those administered by individual universities (as well as the national Center Test) – washback negatively on teaching and learning:

Unfortunately, a number of teachers have raised questions about the validity of the entrance examinations saying that they use out of date testing methods and are mismatched with language teaching curriculum in Japan.

The out-dated testing methods referred to in Brown’s (1995) discussion include discrete-point, multiple-choice grammar questions and translation tasks, both of which, he comments, make up a large percentage of test items. Brown (1995: 277) also identifies the lack of listening and speaking components as two of the main areas of concern regarding curriculum mismatch. The MEXT guidelines place much emphasis on the development of students’ listening and speaking skills, yet an earlier study by Brown and Yamashita (1995) found that in 1993 only 6 out of 21 universities included a listening component as part of their entrance tests and that, a year later, this figure had decreased to 4 out of 21. More recent figures are unknown, although it should be pointed out that, as from 2006, the national Center Test examination will include a listening component. The assessment of speaking skills, however, remains an area of language testing yet to be addressed by individual universities or by the Center Test.

The view that the university entrance examinations in Japan washback negatively on teaching and learning is not shared by everyone, however. For example, Guest (2005), writing in the Daily Yomiuri (28/01/05), claims that:

Time and again it has been shown that university entrance exams do not demand nor focus extensively on English-to-Japanese sentence translation or grammar skills.

Guest (2005) offers no evidence of previous research to back up this claim, although in support of his argument he does provide an analysis of the six sections contained within the 2004 Center Test (see Appendix I) and concludes the following:

1. Examinees are never required to write in Japanese on the test.
2. Examinees are never asked to translate (as an answer).
3. There are no Japanese texts except in the instructions.
4. The skills required correspond to what most highly regarded reading textbooks and teachers demand, especially in terms of meaning-based reading and holistic comprehension.
5. The texts in the test are of a great variety (narratives, dialogues, scientific essays).
6. None of the questions demand...arcane rules or obscure vocabulary.

(Guest 2005 – extracted from the Daily Yomiuri, 28/01/05)

In the same article, it is also suggested that the second stage examinations administered by individual universities ‘are even more comprehensive and holistic in their questions’ (Guest 2005) and with the introduction of listening sections in some cases, are ‘even closer to a four-skill evaluation’ (ibid.). Nevertheless, these observations contradict the research findings of Watanabe (1996) who noted that 94% of national public universities included translation questions in their English examinations – although the figure did drop considerably (46%) for those administered by private universities.

Whilst admitting that ideally, productive skills should be assessed in addition to receptive skills, Guest (2005) foresees two problems with this in relation to the Japanese context. Firstly, he cites the difficulty of administering speaking/listening tests and extended compositions to tens of thousands of examinees which must then be graded in a

short period of time. Secondly, he raises the issue of subjectivity, stating that:

In a system where objectivity is a must (remember that examinees' entire careers may be on the line here) markers cannot afford to let personal, vague or subjective criteria affect the outcomes.

(Guest 2005 – extracted from the Daily Yomiuri, 28/01/05)

Both of these points (particularly the notion that effective subjective marking is almost a non-achievable aim) are contentious issues and bring us back to the comments of Hughes (1989: 47), as noted previously – i.e. we have to think in terms of ‘what will be the cost of not achieving beneficial backwash?’ As Hughes (1989: 47) concludes:

When we compare the cost of the test with the waste of effort and time on the part of teachers and students in activities quite inappropriate to their true learning goals (and in some circumstances, with the potential loss to the national economy of not having more people competent in foreign languages), we are likely to decide that we cannot afford not to introduce a test with a powerful beneficial backwash effect.

The remainder of this study, which reports on the implementation of an original speaking test within the context of a Japanese high school, will adopt a similar line of approach to that of Hughes (1989). Following an introductory section on communicative language testing, the next chapter will proceed with an analysis of the current tests faced by learners in the sample. I would then like to present a rationale for the design of the new test and accompanying marking scheme.

Chapter 3: Approaches to Language Testing

The original test design introduced at the end of this chapter is intended to assess the English speaking skills of Japanese high school students. Whilst it will be used as an end of term assessment in the subject of Oral Communication with a sample group of learners, it is hoped that with some adaptation it could also be administered as part of one of the high stakes tests discussed in section 3.2, below. In order to provide a background to the development of the new test, however, it is important to address two key related areas. Firstly, in view of the fact that the Japanese Ministry of Education has adopted the communicative approach (in theory, if not in practice) as its official methodology, I would like to look briefly at the notion of communicative language testing. Secondly, to offer a greater understanding of the context in which the speaking test will be trialled and thus further clarify the rationale for its design, an analysis of EFL tests currently being administered is also presented.

3.1 Communicative language testing

It is perhaps easier to understand the concept of communicative language testing by comparing it to the teaching and testing methodologies to which it was opposed. Fulcher (2000: 483) notes that the phenomenon initially came about as ‘a reaction against tests constructed of multiple choice items and the perceived over-emphasis of reliability’. Indeed, the discrete-point approach endorsed by Lado (1961) came in for particular criticism (although Fulcher (2000) is careful to note that to a large extent this was unjustified). Morrow (1979: 145), for example, stated that:

An atomistic approach to test design depends utterly on the assumption that knowledge of the elements of a language is equivalent to knowledge of the language... Knowledge of the elements of a language in fact counts for nothing unless the user is able to combine them in new and appropriate ways to meet the linguistic demands of the situation in which he wishes to use the language.

Issues surrounding reliability and validity can be seen as two of the key areas of concern in the development of communicative language tests. The main argument here was that test reliability was often achieved at the expense of validity (and vice versa), with the implication being that the latter was the more desirable of the two. Validity itself, however, can be subdivided into five distinct categories:

1. Face Validity – Does the test ‘look’ reasonable? Do the people who use the test think it is a good test?
2. Predictive Validity – Can the test predict how successful the learners will be at using the language in the future?
3. Content validity – Do the items or tasks in the test match what the test as a whole is supposed to assess? Does the test provide an accurate reflection of the syllabus?
4. Concurrent validity – How do learners’ scores compare with their scores on other language tests? Does the test provide similar results to other tests?
5. Construct validity – Does the test share the same approach as the language programme of which it is part?

* The questions posed above have been adapted from Underhill (1987: 105-108).

Morrow (1979: 147) makes the point that with the exception of face validity and possibly predictive validity, the concept is ‘ultimately circular’:

Starting from a certain set of assumptions about the nature of language and language learning will lead to language tests which are perfectly valid in terms of these assumptions... Thus, a test which perfectly satisfies criteria of content, construct or concurrent validity may nonetheless fail to show in any interesting way how well a candidate can **perform in** or **use** the target language.

(emphasis added)

Davies (1978), cited in Morrow (1979: 146) and Fulcher (2000: 484), referred to the apparent conflict of interest created by the roles of test reliability and test validity as the reliability – validity ‘tension’, defining a test’s validity in terms of the extent to which the

tasks on that test reflect real world activities. Kitao & Kitao (1996) also acknowledge that communicative language tests ‘are judged...on the extent to which they simulate real life communicative situations’:

In testing productive skills, emphasis is placed on appropriateness rather than on ability to form grammatically correct sentences. In testing receptive skills, emphasis is placed on understanding the communicative intent of the speaker or writer rather than on picking out specific details.

(Kitao & Kitao 1996 – retrieved from the World Wide Web)

The above observation by Kitao & Kitao (1996), which notes the importance of producing language that is contextually appropriate, relates to the concept of communicative competence proposed by Hymes (1972) – as cited in Canale & Swain (1980: 4, 7, 15-16). Developed as a response to the competence/performance model put forward by Chomsky (1965) whereby ‘competence’ simply implied a knowledge of grammar, Hymes (1972) introduced the notion of communicative competence which included ‘not only grammatical competence (or implicit and explicit knowledge of the rules of grammar) but also contextual or sociolinguistic competence (knowledge of the rules of language use)’ (Canale & Swain, 1980: 4). Hymes (1972) also made the distinction between communicative competence and communicative performance (i.e. actual language use), with Canale & Swain (1980) noting the implications that this has for teaching and testing within a communicative programme. They state:

...teaching methodology and assessment instruments must be designed so as to address not only communicative competence but also communicative performance, i.e. the actual demonstration of this knowledge in **real** second language situations and for **authentic** communication purposes.

(Canale & Swain 1980: 6, their emphasis)

As we shall see from the following analyses of the tests that are administered within the current context, however, very little consideration appears to have been given to the measurement of actual language performance as it is defined here.

3.2 Current tests of English in the teaching and learning context

The teachers and learners taking part in the study will be introduced in the following chapter. At this point, however, I would like to look at the range of English tests that are currently being administered. It should be noted that the TOEIC and TOEFL tests, whilst taken by some high school students in Japan, are not applicable to any of the learners in the selected sample and are therefore absent from this discussion. Four different categories of test can be identified, however, each of which will now be considered in turn.

Internal end of term tests

Forming part of the school's graduation credit system, end of term tests are administered internally at the beginning of July (1st term), December (2nd term) and March (3rd term). Typically, they tend to focus on language usage as opposed to language use; they employ objective testing methods in preference to subjective methods and they are norm-referenced, rather than criterion-referenced. However, it should be pointed out that as individual teachers are responsible for the design of these tests, there is some scope for change regarding the types of question included.

It could be argued that the end of term tests are, in fact, progress tests – particularly as they take place during the language programme and are administered internally on a reasonably small scale. However, despite taking place three times a year, thus looking back over relatively short periods of learning, I believe that they can be more accurately defined as achievement tests in so far that the objective is to measure 'what has actually

been learned in relation to what is supposed to have been learned' (Ingram 1968: 71). Also, in keeping with the typical characteristics of an achievement test, they provide summative evaluation and are designed to assess rather than assist language learning.

Students need to perform well on the end of term tests in order to accumulate enough credits for their final graduation grade. However, the tests have further implications for those applying to universities via the *suisen* (recommendation) or *ichigei nyushi* (one-skill entrance) procedures – both of which represent alternative paths into higher education. In short, students applying through either of these methods are able to bypass the traditional two-stage university entrance procedure (outlined below) and instead are selected on the basis of information received from high school reports (including internal test results), letters of recommendation, interviews and other criteria. Thus, in addition to assessing how much of the syllabus has been learnt (the main role of an achievement test), end of term tests may also be used for predictive purposes in instances where students are going through the process of a *suisen* or *ichigei nyushi* application. Davies (1968), however, offers a note of caution when using achievement tests for more than one purpose. He makes the following observation:

The use of Achievement tests is to find out how much has been learnt. They are concerned, therefore, entirely with the past though they often are used for predictive purposes. That they should be used predictively, i.e. to make claims about future performance, is probably inevitable though it is not the function of an Achievement test to predict.

(Davies 1968: 7)

Regional tests: the “Seibu-Moshi” and “Kenka-Issei”

About once a month, the learners in the sample are required to sit what I have termed a regional test, of which there are two main types: the *Seibu-Moshi* and the *Kenka-Issei*. Approximately 3,500 students from the 7 high schools in the western district of the

prefecture take the Seibu-Moshi test. Each school in turn assumes the role of ‘central administrator’ and is responsible for writing test papers and collating/reporting scores. Test papers are checked by teachers from each individual school with details of the scores then reported back to the administrating school. The Kenka-Issei follows a similar pattern, but operates on a larger scale with a total of 22 high schools (approximately 14,000 students) from the whole of the prefecture.

Although there is no established “pass” or “fail” grade per se for either test, both can still be regarded as norm-referenced, with the performances of students and schools being compared and then ranked accordingly. This information is not made publicly available, but it is of course provided to each of the participating schools. The testing techniques employed are also similar to those encountered on the end of term tests, with a heavy emphasis placed upon the assessment of reading comprehension and grammar-based questions through the use of objective methods.

Also in common with the end of term tests, it can be observed that neither of the regional tests has a particularly well-defined purpose. Due to the frequency with which they administered, they could be identified as progress achievement tests (Hughes 1989: 12). However, Hughes (1989: 12) stresses the importance of establishing a ‘series of well-defined short term objectives’ on which the tests, the teaching and the syllabus should be based. Unfortunately, such objectives remain unclear in the target situation, while it can also be argued that the issue is further complicated by the existing ‘mismatch’ between language tests in general and the syllabus and recommended teaching practices prescribed by the Japanese Ministry of Education.

The regional tests also exhibit some of the characteristics one might expect to find in a placement test. For example, the information gleaned from both is used to inform

decisions concerning the ‘level’ of university to which a student should consider applying¹. However, whereas students are being grouped in terms of their current language ability, the long-term effects of these tests are clearly of greater significance than the more conventional placement test – the like of which may be administered at the beginning of a language programme.

The STEP / “Eiken” test

The STEP test – also known as the Eiken test – is administered by the Society for Testing English Proficiency, an external testing body that was originally established under the guidance of the Japanese Ministry of Education in 1963. The test is available at seven different levels: 5th, 4th, 3rd, Pre-2nd, 2nd, Pre-1st and 1st, with level 1 being the most difficult. Each level is administered three times a year at external testing centres across Japan, although group applications are also accepted enabling students to sit the test at their own schools.

The STEP test is designed exclusively for Japanese learners of English (with question rubrics provided in the first language) and, as its name suggests, its intended purpose is to measure English language proficiency. Although it is not a compulsory examination, the Ministry of Education states explicitly that high school graduates should on average be at the second level or pre-second level stage of the STEP test (MEXT 2003: 1). However, it is important to note that the STEP test – despite enhancing the future employment prospects of successful examinees – becomes superfluous to the language objectives of many high school students for whom the university entrance examinations assume greater importance. Evidence of student attitudes towards such tests is presented in the research findings of Norris-Holt (2002), who found that the

¹ Information gathered through informal discussions with teaching colleagues suggests that, in the case of the Kenka Issei test (which is taken by approximately 14,000 students), the top 1,500 will be able to make applications to local public universities which, for the most part, are considered more prestigious than private institutions.

Eiken test had greater relevance for junior high school students (12 – 15 year olds) than for high school students.

As with the previously mentioned tests, the Eiken is designed to assess rather than assist language learning and, in order to achieve high reliability, it adopts objective testing techniques. In fact, all items are multiple-choice enabling answer papers to be machine-marked – the obvious disadvantage being that the indirect assessment of writing could result in negative washback. A direct test of speaking is available, with ‘second stage’ interview tests being administered to examinees taking the 3rd grade level test and above.

The fact that successful examinees of the STEP test fall on or above a cut-off score and that the average scores of each section are reported to all candidates (thus offering a point of comparison), undoubtedly suggests that it is norm-referenced. However, the performance descriptors that are provided for each level (see Appendix II) also point to criterion referencing – a fact that has led MacGregor (1997: 28) to refer to the STEP test as a ‘hybrid’ criterion-referenced / norm-referenced language test.

University entrance examinations

Of all the English tests taken by high school students in Japan, those administered as part of the university entrance system are often regarded as the most important, representing the ‘primary, or even sole, purpose’ of language study for many learners (Brown 1995: 272). Watanabe (2004: 131) offers further clarification of the situation:

Because there is no school-leaving exam, and because high school records are not given much weight, the entrance examinations are likely to be very high-stakes for the applicants.

Universities in Japan fall into one of three categories: national (managed by central government), local public (managed by local governments) and private². In the case of national and local public institutions, the entrance procedure consists of two stages. The first stage examination, designed by the National Center for University Entrance Examinations (NCUEE) and known as the Center Test, is administered nationally at designated testing centres each year at the end of January. Although the vocabulary used on the Center Test is limited to the “approved” word list put forward by the Ministry of Education, it is not the purpose of the test to look back over what has been taught in high school English classes. Instead, it is essentially a norm-referenced placement test and therefore forward-looking. For example, it is not a test with a “pass” or “fail” standard, but rather one in which an examinee’s score and final placing will determine the ‘rank’ of university to which he or she can apply. Thus, examinees in the top percentile are able to apply for places at the more prestigious universities.

This represents the point at which students enter the second stage of the entrance procedure when they go to sit the individual examination administered by their chosen university. It should be noted, however, that the entrance requirements of private universities differ slightly to national and local public institutions, with most simply using their own examination to select students. This single examination selection process is also preferred by junior colleges offering two-year courses³.

The entrance examination system is further complicated by the fact that most universities – whether national, local public or private – will often have more than one English test as each department produces its own paper. With over 700 universities in

² Official Ministry of Education figures published in 2003 put the total number of **universities** in Japan at 702, of which 100 were national institutions, 76 were local public institutions and 526 were private (retrieved from the MEXT website: www.mext.go.jp/english/statist/index.htm).

³ Official Ministry of Education figures published in 2003 put the total number of **junior colleges** in Japan at 525, of which 13 were national institutions, 49 were local public institutions and 463 were private (extracted from the MEXT website: www.mext.go.jp/english/statist/index.htm).

Japan, a vast number of entrance tests are therefore administered each year. It is perhaps not surprising, then, to find that the content of those tests also varies somewhat. For example, the first stage Center Test – which was discussed in some detail at the end of the previous chapter – consists entirely of objective items designed to assess vocabulary, grammar, pronunciation and reading. The second stage examinations, administered by each individual university department differ again:

As if making up for the NCUEE, the second-stage examinations of national/local public universities consist of subjective items, including translation of English into Japanese, Japanese into English, writing, and so forth. The private university examinations exhibit even greater variety in their test contents and methods; some examinations include grammar, vocabulary, and reading, whereas others also include listening and writing. The items of these tests may consist of cloze, short-answer, multiple-choice, picture description, and a number of other varieties.

(Watanabe 2004: 132)

Through a comprehensive analysis of 21 university entrance examinations administered in 1994, Brown & Yamashita (1995) confirm the extent of this variation in test content and methodology. The study included the Center Test along with the test papers of 10 private, 8 national, and 2 local public universities – the latter two being grouped together under the single heading of ‘public’ institutions. In order to determine whether the examination content varied from year to year, a comparison of the 1994 and 1993 examinations was also offered.

Whilst it is beyond the scope of this discussion to analyse the findings of the Brown & Yamashita (1995) study in specific detail, it is worth mentioning some of the more salient features. Of particular interest, with regard to the washback effect of these examinations, is the nature and balance of the test items employed by the different types of institution. Indeed, a great deal of diversity was found to exist with private universities generally relying more upon multiple-choice items and public universities placing a

heavier emphasis on short answer/essay and translation questions. The Center Test consisted almost entirely of multiple-choice items. Thus, questions in which examinees are actually required to produce language are more prevalent in the public university test papers. However, further qualification is required here, with Brown & Yamashita (1995: 95) noting that:

...while the public university examinations are laudably using more integrative items, they are also relying heavily on translation items.

In a study conducted the following year, Watanabe (1996) also found that public universities tended to favour test items requiring some form of translation. Watanabe (1996) reports that out of the most popular 280 university department tests, 63% included translation items. However, in terms of the types of institution using translation items, 94% of the tests came from national or local public departments as opposed to 46% from private university departments.

Whether the entrance examinations are set by national, local public or private universities, the possibility of negative washback on teaching pedagogy needs to be acknowledged, particularly when there is such reliance upon discrete-point, multiple-choice items (in the case of the Center Test and private universities) and translation tasks (in the case of national and local public universities).

3.3 Rationale for the new speaking test

Aims of the test

In order for the communicative syllabus aims of the Japanese Ministry of Education to be implemented effectively, it is important to develop tests that more accurately reflect those aims. As illustrated in section 3.2, a major drawback of the examinations currently taken by Japanese EFL learners is that they involve very little – if any – direct testing of either of the productive skills of writing or speaking. In order to address this issue, an original

speaking test has been designed here for trial with a group of pre-intermediate learners (see Appendix III). It should perhaps be noted, however, that whilst I have categorised the learners as ‘pre-intermediate’, their speaking ability remains at a relatively low level.

It is important to note that throughout the design of the test, consideration obviously had to be given to the learning context in which the actual trial will take place. Thus, certain aspects of the design rationale are unavoidably context-specific, although – as mentioned earlier – it is of course hoped that the test could be adapted for use in other situations. As it will serve as an end of term assessment in the context of this particular trial, it is essentially being used here as an achievement test. Part two of the test in particular (giving directions) reflects an area of the syllabus covered in language classes during the previous term.

Test type

The arrangement of the test (as illustrated in figure 3.1 on page 41) allows for two learners who, for the most part, will interact with each other. As noted by Brown & Yule (1983: 111) and Underhill (1987: 30) there are certain advantages to be gained from pairing learners together. For example, it is easier to create the genuine information gap necessary for communicative interaction. There is greater equality between the speakers, which not only ‘makes them more willing to speak’, but also ‘makes the communication as fluent and as successful as possible’ (Underhill 1987: 30). As a result, the paired format arguably encourages a more natural interaction; the following example provides an illustration of the type of problem that can arise through the ‘unequal’ teacher – student interaction:

The established status of the teacher is that he knows more than the student about most things encountered in the learning situation. If the teacher gives the student an object to describe, then the student has to create, for himself, an

artificial information gap between his knowledge and the teacher's. He has to behave as if the teacher doesn't know what the object looks like. Why should we require this additional, and highly artificial, dimension in the student's behaviour, when it is only his ability to use spoken English which we wish to assess?

(Brown & Yule 1983: 111)

For trial purposes the interlocutor also acts as the assessor, although to counter the possible loss of reliability that might arise through one person carrying out this dual role, the test is taped for the purpose of re-marking.

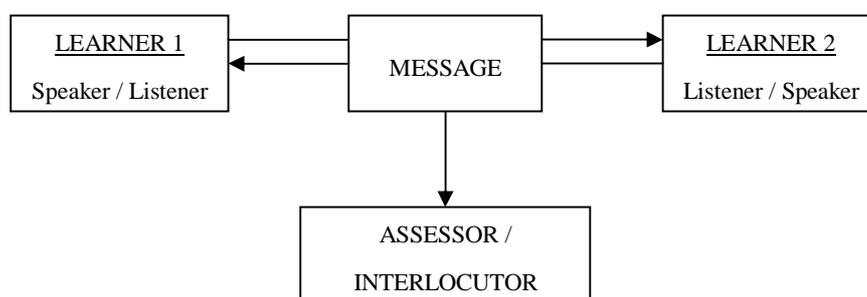


Figure 3.1 The pattern of interaction for the proposed speaking test (adapted from Underhill, 1987: 29)

The decision to adopt a one-on-two format was also influenced by two other factors. Firstly, the limitations of the trial context had to be considered, particularly with regards to the availability of teachers who would have enough time to fulfil the role of either interlocutor or assessor. Secondly, in view of the fact that the test has been designed to serve as an end of term internal assessment for a specific group of learners, it seemed more logical that the role of assessor/interlocutor should be performed by the teacher who works regularly with those learners.

For the test to be adapted for use in other contexts, any decision around whether or not to include two examiners – one as an interlocutor, the other as an assessor – would really depend upon the financial and time constraints of the situation. Whilst the inclusion of two examiners would allow reliability to be monitored more effectively, the

obvious advantage of the current format is that any assessor training would be more cost effective, as fewer people are required to administer the test.

Elicitation techniques

The test consists of three parts, which are outlined as follows:

Part one

The two learners briefly introduce each other to the assessor/interlocutor, who then also interacts with each candidate by asking further questions based around topics such as family, school life, travel⁴ and future plans. Thus, during this opening phase of the test, which can be seen as a ‘warm-up’ to put the learners at ease, assessment will focus upon interactional short turns of speech.

Part two

In part two, each learner is given a simple map and is asked to direct his/her partner from one place to another. It is a guided speaking task with an ‘information gap’, and is designed to elicit a longer turn of speech. This time, the language produced will be transactional. The fact that the task is highly structured also makes it a suitable activity bearing in mind the language ability of the learners involved in the trial.

Part three

Part three is a ‘spot the difference’ activity involving two pictures and has been adapted from an idea by Brown & Yule (1983: 123-126). It is essentially a descriptive task, but in order to meet the communicative requirements of the test, there is the added element of problem-solving. As the two learners negotiate the solution to the problem, it will be necessary for them to alternate between transactional, long and short turns of speech.

⁴ All learners in the sample have experienced a two-week homestay programme in the UK, hence the inclusion of the question, ‘Have you ever visited any other countries?’ (see Appendix III: Speaking test, part one, option card 2).

It is hoped that the inclusion of information-gap activities will have a favourable washback effect upon learning. At present, there is a tendency for some learners to rely too heavily upon their first language when engaged in similar tasks in the classroom. However, they have also been included in order to enhance the communicative value of the test and were considered more appropriate than role-plays (another common classroom activity) for the assessment of spoken language skills. Indeed, as a classroom activity, role-plays often meet with limited success due to the reluctance of learners to assume the part of another person. There is also, of course, no information-gap to motivate the learners when using a role-play. Although one could argue that there is an element of role-playing in part two of the test in which the listener must play the part of a tourist asking for directions, the existence of an information gap at least gives the task a degree of process authenticity. Also, whilst it is difficult to think of a scenario in Japan in which the learner will have to use English in order to ask for directions, the situation could arise should he or she travel abroad in the future.

Timing of the test

It is estimated that the test will take 10 – 12 minutes per pair of learners, with part one lasting for 4 – 5 minutes, part two for 3 – 4 minutes and part three for 3 minutes. Although Hughes (1989: 105) recommends a duration time of 30 minutes, arguing that ‘it is unlikely that much reliable information can be obtained in less than about 15 minutes’, a longer test in the trial context would be difficult to administer as assessment could only take place during regular timetabled classes. A short oral test, such as the one proposed, could also be applied more easily to a high-stakes context, such as the regional or university entrance examinations discussed in section 3.2.

Marking system

An analytic performance scale (see Appendix IV) is used to score the test with learners assessed according to the following criteria: pronunciation, vocabulary and grammar, fluency and spontaneity, and communication. Each of these criteria has 5 different score bands: 0, 1 – 2 points, 3 – 4 points, 5 – 6 points and 7 – 8 points. Thus, a maximum of 8 points can be scored in each of the four marking categories, meaning that the test is out of a total of 32 points. In this particular learning context, the final scores are made into percentages so that they can be applied to the 5-point grading scale adopted by the school, whereby at the end of term, students receive a score of between one and five on their report cards for each subject. This is usually calculated according to their performance in the end of term tests, as follows:

Table 3.1 The 5-point grading scale

TEST SCORE %	0 – 29	30 – 44	45 – 69	70 – 84	85 – 100
TERM GRADE	1	2	3	4	5

As mentioned earlier, the new speaking test is taped to allow for re-checking. Depending on the context in which it is used, this would ideally be carried out by a second assessor.

The actual trial of the speaking test will be discussed in the following chapter. After introducing the teachers and learners involved in the trial, I will explain how the washback effect of the existing tests and the new test is measured through classroom observation and teacher/student questionnaires.

Chapter 4: Test Washback Within the Sample Teaching and Learning

Context

This chapter is concerned primarily with the research procedure – the actual results of the study will be presented in the following final chapter. In this part of the discussion I will provide a brief outline of how the new speaking test was trialled with a sample group of learners (see section 4.3). First of all, though, it is important to remember that another central aim of this study is to examine the washback effects of the current range of English tests being taken by the learners in the sample context. As already noted, whilst it is often argued that many of these tests have a detrimental effect on teaching and learning, much of the research that has been conducted to date is divided over the nature and extent of this washback. Classroom observation and teacher and student questionnaires were used to explore this issue further – the details of which are outlined in section 4.2. The washback effect of the speaking test was again measured by administering a questionnaire to the students involved, whilst discussions with teachers as well as the opportunity to observe the learners firsthand during timetabled classes in the weeks prior to the test, offered more informal methods of data collection.

The decision to include classroom observation and questionnaires as complementary methods of data collection is of particular relevance:

If the core of any definition of washback has to do with the effects of tests on learning and teaching, then it is necessary to document those effects – both by **asking about** and by **watching** teaching and learning.

(Bailey 1999: 36 – emphasis added)

Bailey (1999: 37) elaborates further by citing Wall & Alderson (1993: 65) who made the point that by observing classes we are able to discover the amount of impact that a test has on methodology, but of equal importance – and to understand why that test affects teaching and learning in the way it does – it is also necessary to speak to the participants

involved in the observation session. This could be via interview or discussion (as was the case with the Wall & Alderson study), or primarily through questionnaires (as in this study).

Even with a relatively small scale study such as this, however, another advantage of using observational evidence is that to some extent it allows us to confirm the accuracy and reliability of the questionnaire data, or ‘to see whether what teachers and learners say they do is reflected in their behaviour’ (Alderson & Wall 1993: 127).

4.1 The teaching and learning context

The sample consisted of teachers and learners taken from the upper secondary education (i.e. high school) sector in Japan. Most of the research was conducted at one participating school – a private high school of 486 students in the south of Japan. However, additional data was also collected (via questionnaire) from teachers working at public high schools in the area.

In line with the Japanese educational system, the participating high school has three year grades. Typically, following three years at junior high school, students enter the first grade at sixteen years of age and graduate at the age of eighteen at the end of the third grade. Each grade consists of six homeroom classes in which students remain for the study of all subjects on the curriculum. Whilst for the most part this serves to create “mixed-ability” groups, each grade does also have two streamed classes – one of which is aimed specifically at learners hoping to move on to university after graduating from high school. For this particular group, English study is divided into three separate subject areas: reading (English IR and English IIR), grammar (English IG and English IIG) and Oral Communication (OCI and OCII). All other classes follow a similar pattern for Oral

Communication but, in the case of English, reading and grammar are usually combined into a more general course of study (English I and English II)¹.

4.1.1 Teachers in the sample

For the classroom observation sessions, two Japanese teachers of English agreed to take part in the study. In terms of experience, teacher A – who is in his early 30's – graduated from a private university and has held his current position for three years. He has 19 timetabled lessons a week, most of which involve teaching reading ('English R') and grammar ('English G') classes on the upper level course. Teacher A has taught English for a total of nine years and previously worked at a private junior high school.

Teacher B is in his early fifties and graduated from a public university. He has 28 years experience and currently teaches 18 lessons a week, mainly on the English I and English II courses, with just one upper level class for English IG. Teacher B also runs his own private cram school (or juku), offering additional evening and weekend study for junior high and high school students.

In addition to classroom observation, teacher questionnaires were also administered in an attempt to determine the effect that current tests have on teaching content and methodology. As well as being given to the two teachers involved in the observation sessions, the questionnaire was also completed by a further five high school teachers in order to extract a wider sample of responses and increase the validity of the research: two were working at the same private high school as teachers A and B, whereas the other three were employed at public high schools in the area.

¹ English I, IR, IG and OCI represent first grade courses, whilst English II, IIR, IIG and OCII are followed during the second and third grades.

4.1.2 Learners in the sample

A total of six groups of learners participated in the study – two 1st grade classes, two 2nd grade classes and two 3rd grade classes. The two first grade classes and one of the second grade classes served to provide the data collected from the observation sessions. The remaining second grade class and the two third grade classes provided the data collected from the student questionnaires and were also involved in the trial of the speaking test. Further details of the characteristics of these groups – for example, class size and course details – will be provided at the relevant points in the proceeding discussion.

4.2 Measuring the washback effect of current EFL tests in the sample context

4.2.1 Classroom observation

The observation sessions were carried out to address in part the first research question which, to recapitulate, speculated on the extent to which teachers were influenced by test content. The amount of communicative methodology (the preferred approach of MEXT) that teachers actually implemented at classroom level was also relevant to this area of the study.

There were four observation sessions in total – two involving teacher A, and two involving teacher B. In the case of teacher A, data was collected from two 1st grade groups of learners – one an English IR (i.e. reading) class, the other a general English class. The first session, which was scheduled as a 45-minute lesson, involved a group of just 8 learners following a course of study aimed specifically at students hoping to gain entry into higher education. The second session was scheduled for 50 minutes and consisted of a group of 14 learners, who hope to enter full-time employment or alternatively, enrol on a two-year course at a junior college after graduating from high school.

With teacher B, data was collected from two 45-minute lessons, both consisting of groups of learners studying on the general English course. The first of teacher B's sessions involved the same group of learners that was observed during the general English class with teacher A above. The second session included a mixed-ability group made up of 27 second grade students.

The above information is summarised in table 4.1 below, with the end column indicating the on-going examinations that each group will take at various points over the next two years. It is also reasonable to assume that a small percentage of students from all three groups will opt to take the non-compulsory STEP proficiency test during their time at high school.

Table 4.1 Outline of the groups involved in the observation sessions

Session No.	Group No.	Teacher	Time	Year	No. of students	Course of study	Examinations
1	1	A	45 mins	1 st grade (streamed)	8	English IR	Seibu-moshi / Kenka-Issei End of term tests (University entrance exams)
2	2	A	50 mins	1 st grade (streamed)	14	English I	End of term tests (Junior College exams?)
3	2	B	45 mins	1 st grade (streamed)	14	English I	End of term tests (Junior College exams?)
4	3	B	45 mins	2 nd grade (mixed-ability)	27	English II	End of term tests

Data from all four sessions was collected and analysed using Part A of the Communicative Orientation of Language Teaching Observation Scheme (COLT) (Spada & Frohlich 1995). Specifically, Part A of the scheme provides us with a useful framework with which we can observe – in ‘real time’ – the teaching and learning that takes place in second language classrooms. It should be noted, however, that the sessions

were also taped in order to verify certain aspects of the data later on. The COLT scheme consists of five main features²:

- Participant organisation: Within this area of the scheme, details of the way in which students are organised during a classroom activity are recorded. This includes categories such as teacher to student or class ($T \leftarrow \rightarrow S/C$), whereby the teacher interacts with the whole class and/or individual students. This is the type of interaction that occurs frequently in traditional teacher-centred classrooms. Other patterns of interaction included within this feature are group tasks and individual tasks.
- Content: This feature ‘refers to the subject matter/theme of activities; that is, what the teacher and students are talking, reading or writing about or what they are listening to.’ (Spada & Frohlich 1995: 16). Perhaps most importantly within the context of this study (bearing in mind the apparent conflict between the communicative teaching aims of MEXT and the nature of English examinations) it was possible to determine the extent to which the focus of instruction centred around meaning as opposed to form.
- Content control: There are three categories within this field – teacher/text, teacher/text/student and student. According to these categories then, the recorded data indicates who selected and controlled the tasks that were carried out during each observed class.
- Student modality: This area of the scheme focuses entirely on the students and is used to record whether they are listening, speaking, reading or writing during a classroom activity, or whether they are engaged in a combination of these skills at any one point.
- Materials: The type and source of materials that are being used during a classroom activity are recorded under this feature. Among the possibilities are minimal texts

² An example of a COLT observation sheet can be found in Appendix V.

(e.g. isolated sentences, word lists) or extended texts (e.g. stories, dialogues, paragraphs). The source of the materials is classified as authentic, non-authentic, adapted or student-made.

All five features are coded within the context of ‘activities’ or ‘episodes’ which make up the basic units of analysis for the scheme. Activities and episodes are defined as ‘separate units which constitute the instructional segments of a classroom’ (Spada & Frohlich 1995: 14):

Separate activities would include such things as a drill, a translation task, a discussion or a game.

Three episodes of **one** activity would be: teacher introduces dialogue, teacher reads dialogue aloud, individual students read parts of dialogue aloud.

(Spada & Frohlich 1995: 14)

During the classroom observation sessions, the starting time and a brief explanation of each activity/episode was recorded. A check mark was then placed on the observation sheet (see Appendix V) against any of the COLT categories that occurred within those episodes and activities. For example, how was the class organised? Were the students working in groups or individually? What was the main pattern of interaction? (Was the teacher leading one central activity and interacting with the whole class or were the students perhaps involved in choral work?) Which of the four language skills were being utilised?

During any one activity or episode it was of course possible for one or several categories to be checked off. When just one category is checked, it is referred to as having exclusive focus. When a combination of categories is checked, it can have either primary focus or equal focus:

A primary focus is when most of the time is spent on a particular category. An equal focus is when approximately the same amount of time and emphasis is spent on more

than one category.

(Spada & Frohlich 1995: 31)

After the observation sessions had ended, the amount of time (in minutes) that had been spent on each category of the scheme during each activity/episode was calculated. The total amount of time for each of the categories (under the main features of Participant, Organisation, Content, Content Control, Student Modality and Materials) was then calculated and recorded as a percentage of the total class time. It should be noted that, in accordance with the conventions of COLT, any categories that appeared in combination that **did not** have exclusive, primary or equal focus, were ignored during these final calculations. Also, when categories occurred in combination, they were calculated separately and subsequently recorded separately. So, for example, when students were involved in an activity that required listening and speaking, this was reported as a single category (under the feature of Student Modality) in the final table of results (see Appendix VI). An analysis and discussion of the results can be found in the following chapter.

4.2.2 The survey instrument: teacher and student questionnaires

In addition to classroom observation, teacher and student questionnaires (copies of which can be found in Appendices VII and VIII) were also administered in order to measure the washback effect of EFL tests currently taken in the sample context. The data collected from the teacher questionnaires (which, as noted earlier, were administered to a total of seven high school teachers) supplemented the data from the observation sessions and were used to determine the effect that the tests have on teachers and teaching.

The student questionnaires, on the other hand, were of course more concerned with the effect that the tests have on learners and learning. These were administered to two third grade classes and one second grade class. However, as the second grade class was

following an identical course to one of the third grade classes and both sets of learners shared similar objectives (i.e. they will take university entrance examinations in the near future), these two groups were combined for the purposes of recording and reporting data – see table 4.2 below.

Table 4.2 Outline of the groups involved in the questionnaire survey

Group No.	Year	No. of students	Courses of study	Examinations	Test Trial
4	3 rd grade (streamed) & 2 nd grade (streamed)	27	English IIR English IIG Oral Communication II	Seibu-moshi / Kenka-Issei End of term tests University entrance exams 2006 & 2007.	Yes
5	3 rd grade (streamed)	20	English II Oral Communication II	End of term tests Junior College exams 2006	No

As can be seen from the table, the learners in groups 4 and 5 are in streamed classes. However, the students in group 5 are unlikely to sit university entrance examinations and, therefore, the survey data for this group were recorded separately to enable comparison with the set of data collected from group 4. It can also be noted from the last column of table 4.2 that the new speaking test was not trialled with group 5.

Although both questionnaires were quite lengthy, they consisted mainly of Likert-scale questions in which the respondents were asked to record their opinions by circling a numbered category from 1 – 4 (e.g. 1=strongly disagree; 4=strongly agree). This meant they could be completed and returned in a relatively short space of time. Additional space was also provided after each question for any further comment or clarification, however.

Whilst Likert questionnaires often use a response scale of 1 to 5, it was felt that in this particular context the 1 to 4 scale would be more effective in order to discourage

respondents from relying too heavily on the neutral or “non-opinion option” that exists in the former. This follows a suggestion made by Brown (2001), who has encountered similar difficulties when using Likert-scale questions on survey instruments:

Some types of students tend to “sit on the fence” on Likert questions.. If you need to force respondents to express a definite opinion, that is, to jump one way or the other, on each question, you may find it useful to have an even number of options from which they must choose...

(Brown 2001: 41)

One final point needs to be made about the choice of language for each questionnaire. In the case of the teacher questionnaires, all questions and question rubrics were presented in English. Due to the language level of the learners in the sample, however, the questions and question rubrics on the learner questionnaires were presented in Japanese and English, the latter being included so that the data could be recorded more easily. The results of the questionnaires are analysed using basic descriptive statistics in the final chapter.

4.3 Trialling the speaking test and measuring washback

The speaking test was trialled with group 4 (n=27) during the final two weeks of June, with students’ scores from the test acting as their final end of term grade in the subject of Oral Communication. As discussed in the previous chapter, it was administered to learners in pairs, with one interlocutor/assessor present. All learners were informed about the test at the beginning of the school term (i.e. the first week in April) and were provided with information regarding the test format and mark scheme. During Oral Communication classes in the weeks leading up to the final assessment, students also practised speaking tasks which were similar in style to the test items. This process accorded with Hughes’s (1989: 46) view that it is particularly important that information

‘should be made available to everyone concerned’ when introducing a new test with new testing methods (see chapter 2, page 25).

The washback effect of the speaking test was determined via two methods. Firstly, a second questionnaire (see Appendix IX) was administered to those learners involved in the trial, not only to gauge their reactions to the test itself, but perhaps more importantly to find out whether they felt that it had improved their learning in any way. It was also my intention to discover whether or not the presence of a speaking test had acted as a motivational factor and whether the learners had prepared any differently for this particular test. The results of this questionnaire are also analysed statistically in the final chapter.

Secondly, qualitative data was also used in order to establish whether the introduction of the new test had had the desired (positive) effect on learning. Throughout the weeks leading up to the trial of the test, I was able to engage in a certain amount of ‘reflective’ classroom research, observing any changes in learning behaviour. Following the trial, the test was also shown to teachers A and B, who provided some useful informal feedback – again, their comments will be included in the following chapter.

Chapter 5: Research findings and implications

The results of the study are presented in this final chapter, which is divided into three main areas. The data collected from the classroom observation sessions are considered first, followed by the data from the teacher questionnaires, and then finally the data from the two student questionnaires. In conclusion, I would like to focus on the implications the results have for the design of EFL tests and for the successful implementation of communicative methodology in the Japanese context.

5.1 Classroom observation results and discussion

As noted previously, the data were collected using the COLT (Part A) Observation Scheme (Spada & Frohlich 1995). To recapitulate, the sessions were carried out in order to observe the teaching practice of the teachers in the sample, not simply in terms of what was being taught (i.e. content), but also in terms of how it was being taught (i.e. methodology). Of particular interest, in relation to the first research question, was the extent to which current examinations might influence teaching methodology. In other words, the results of the observations were primarily concerned with the fourth of Alderson & Wall's washback hypotheses, i.e. that 'a test will influence **how** teachers teach' (Alderson & Wall 1993: 120). In view of what has already been said about the nature of the examinations taken by the learners in this sample, there was a strong likelihood that as a result of test washback, the two teachers would to some extent rely on grammar-translation (GT) – at the expense of a more communicative approach – in all four of the observed classes. However, bearing in mind the different examination goals of each group (i.e. group 1: university entrance examinations; group 2: possibly junior college examinations and group 3: internal end of term tests) it was predicted that:

1. There would be a higher incidence of grammar-translation in session 1 (Teacher A/Group 1).
2. There would be a lower incidence of grammar-translation in session 4 (Teacher

B/Group 3).

3. Similar teaching methodology would be observed in sessions 2 and 3 (Teacher A/Group 2; Teacher B/Group 2).

Although touched upon briefly in chapter 1, it is perhaps important at this stage to clarify what is meant by the grammar-translation approach in order to highlight the patterns of behaviour that we might expect to emerge in a typical GT-orientated class. Richards and Rodgers (1986) offer a summary of the main characteristics:

1. Grammar translation is a way of studying a language that approaches the language first through detailed analysis of its grammar rules, followed by application of this knowledge to the task of translating sentences and texts into and out of the target language. It hence views language learning as consisting of little more than memorizing rules and facts in order to understand and manipulate the morphology and syntax of the foreign language.
2. Reading and writing are the major focus; little or no systematic attention is paid to speaking or listening.
3. Vocabulary selection is based solely on the reading texts used, and words are taught through bilingual word lists, dictionary study, and memorization.
4. The sentence is the basic unit of teaching and language practice.
5. Accuracy is emphasized.
6. Grammar is taught deductively – that is, by presentation and study of grammar rules, which are then practiced through translation exercises.
7. The student's native language is the medium of instruction.

(Taken from Richards & Rodgers 1986: 3-4)

The full table of results for the four classroom observation sessions is provided in Appendix VI. For the purposes of the following discussion, the more salient features are

analysed and reproduced in tables 5.1 – 5.4. All recorded figures represent the total percentage of class time spent on each category of the COLT scheme under the main features of participant organisation, content, content control, student modality and materials.

Let us begin with participant organisation, which was ‘developed to describe distinctions between teacher-centred and group-work interactions in L2 classrooms’ (Spada & Frohlich 1995: 15). Within this feature it can be noted that for the majority of the time during both sessions, teacher A was engaged in whole-class interaction (T<->S/C). However, in the case of session 2 the approach was almost exclusively lecture-based with significantly more time (94%) being spent interacting with individual students or the whole class here than in session 1 (63.41%). When teacher A was not engaged in whole-class interaction, students spent the rest of the time working individually – see table 5.1 below.

Table 5.1 Patterns of participant organisation adopted by teacher A (% of total class time)

	T<->S/C	Individual
Session 1	63.41	36.59
Session 2	94.00	6.00

Note: T<->S/C: Teacher to individual student or whole class.

Teacher B also favoured a teacher-centred approach to learning, but did adopt a wider variety of techniques during session 3, as illustrated in table 5.2.

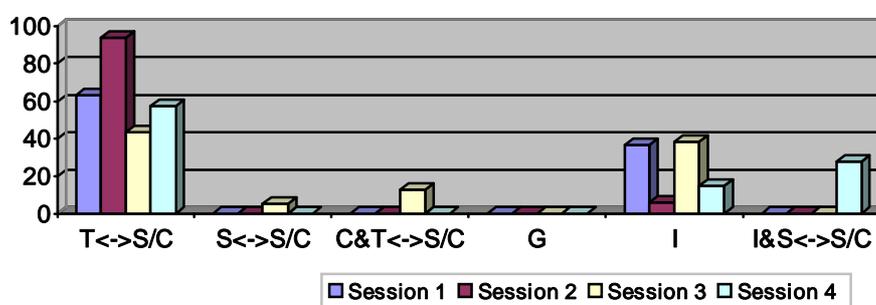
Table 5.2 Patterns of participant organisation adopted by teacher B (% of total class time)

	T<->S/C	S<->S/C	Choral & T<->S/C	Individual	Individual & S<->S/C
Session 3	43.59	5.13	12.82	38.46	0.00
Session 4	57.50	0.00	0.00	15.00	27.50

Notes: T<->S/C=Teacher to individual student or whole class; S<->S/C=Student to student or student to whole class.

Presented from a different angle, the above data can also be considered by way of the following diagram (figure 5.1).

Figure 5.1 Patterns of participant organisation during the four observation sessions



Notes: T<->S/C=Teacher-led activity; S<->S/C=Student-led activity; C&T<->S/C=Choral work within teacher-led activity; G=Group work; I=Individual work; I&S<->S/C= Individual work/Student-led activity.

The dominant role assumed by teacher A during session 2 (T<->S/C=94%) perhaps suggests that, contrary to expectations, this was in fact the least communicatively-orientated class. However, it should also be noted that for the most part, the high incidence of individual work recorded during session 1 (Teacher A/Group 1), and during session 3 (Teacher B/Group 2), involved written translation tasks (at sentence level). Whilst students were working on these written tasks, both teachers tended to interject at regular intervals (in Japanese) to check on learners' progress. Thus, it is suggested that in terms of participant organisation, the teacher-centred classroom – which arguably allows teachers to implement the grammar-translation method to the greatest effect – was prevalent throughout all four sessions. Certainly, figure 5.1 confirms that teachers spent a

lot of time engaged in teacher-led activities (T<->S/C), whilst the absence of group work offers further evidence of a teacher-centred, grammar-translation approach.

Indeed, there was little, if any, opportunity for students to interact with one another. The two incidences of “student-led” activities reported in sessions 3 and 4, the latter occurring during a writing task, amounted to an error correction exercise, with individual students being selected to write their answers on the blackboard. Thus, the group checked their work together and all students worked at the same pace in a classic “lockstep” learning environment.

The next feature of the COLT scheme is content, which allows us to determine the extent to which a lesson focuses on meaning or form. In the case of all four sessions in this study, the focus of instruction was exclusively form-based, i.e. 100% of the time in each class was spent on discrete-point sentence exercises based around grammar and vocabulary. Classes were also conducted in the first language – indeed, this applied to initial greetings and procedural directives as well. Perhaps not surprisingly given the preferred teaching approach, a figure of 100% was also recorded under the category of teacher/text in the following feature of content control, indicating that students played no part in the selection of the topics or tasks.

The final two features – student modality and materials – also offer a strong indication that grammar-translation remains the dominant methodology within the Japanese EFL classroom, regardless of any specific examination target. In teacher A’s classes, the amount of time devoted to each of the four language skills is illustrated in table 5.3.

Table 5.3 Student modality during teacher A's classes (% of total class time)

	Listening	Listening & Reading	Listening & Writing	Writing
Session 1	0.00	9.76	53.66	36.59
Session 2	20.00	46.00	22.00	12.00

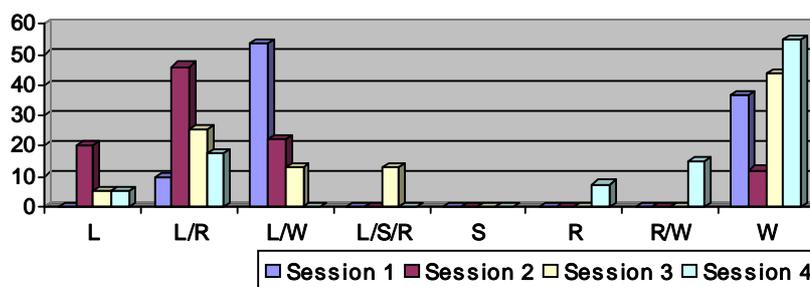
Teacher B, on the other hand, devoted more time to writing tasks and, as was the case within the feature of participant organisation, appeared to adopt a little more variety in his approach.

Table 5.4 Student modality during teacher B's classes (% of total class time)

	Listening	Listening & Reading	Listening & Writing	Listening Speaking & Reading	Reading	Reading & Writing	Writing
Session 3	5.13	25.64	12.82	12.82	0.00	0.00	43.59
Session 4	5.00	17.50	0.00	0.00	7.50	15.00	55.00

Once again, it is useful to consider these two sets of data by way of graphical representation (see figure 5.2, below).

Figure 5.2: Comparison of student modality during the four observation sessions



Notes: L=Listening; L/R=Listening & Reading; L/W=Listening & Writing; L/S/R= Listening, Speaking & Reading; S=Speaking; R=Reading; R/W=Reading & Writing; W=Writing.

Again, further clarification is required with certain aspects of both sets of data, particularly in view of the high percentage of time that students were engaged in listening (either as a discrete skill or in combination with reading and writing). It should be remembered that as lessons were conducted in Japanese, any incidences of listening that were observed did not constitute L2 listening activities as such.

The percentage of time spent reading and writing also needs further qualification as the materials being used were designed specifically for Japanese EFL students and perhaps more importantly, were classified (and recorded) as minimal texts (i.e. isolated sentences, word lists). During sessions 1 and 2, teacher A used copied handouts taken from a grammar textbook. Despite the different aims of each group (i.e. group 1 were preparing for the regional tests and – in the longer term – university tests, whereas group 2's needs focused mainly on internal end of term tests), both were given identical materials taken from the same source. Tasks included direct translation of sentences (L1 to L2 and L2 to L1); tag question/sentence completion exercises; and identifying/translating adverbs at sentence level. During sessions 3 and 4, teacher B used self-prepared handouts containing translation exercises – again, both L1 to L2 and L2 to L1.

Another telling feature of both sets of data within these two features is the absence – across all four sessions – of L2 speaking (or activities which facilitate speaking). This suggests that, in the case of this sample context at least, teachers are at present struggling to implement the official communicative aims of MEXT at classroom level.

It is extremely difficult, then, to find evidence in support of the initial prediction that the amount of grammar-translation methodology employed in each session would vary according to the type of examination being administered to each group of learners. This

is due mainly to the fact that all four sessions tended to epitomise the grammar-translation approach to learning and indeed, the seven characteristics identified by Richards & Rodgers (1986: 3-4) could certainly be found in each one. The fact that group 1 was preparing for university tests and group 3 for internal end of term tests appeared to make little difference in terms of how English was actually taught. As this formed the basic premise of the prediction that session 1 would have a higher incidence of grammar-translation than session 4, it remains uncertain whether changes to examinations would necessarily bring about positive changes to teaching methodology.

The following area of the research findings, which analyses the data collected from the teacher questionnaires, may help to shed more light on this area of debate.

5.2 Survey results and discussion

5.2.1 Teacher questionnaire

As noted in the previous chapter, the teacher questionnaire (see Appendix VII) was administered to seven Japanese high school teachers of English. It was designed to measure teachers' attitudes towards English education in Japan and of course, to establish the extent to which current EFL tests influence their teaching approach. Although made up of a total of 16 questions, numbers 4 to 14 (which were based on a 4-point Likert scale) consisted of multiple parts – tables 5.5 and 5.6 (pp.65-66) show the responses of the seven teachers to these particular questions. In order to clarify the method of data analysis I shall, at this point, offer a brief explanation of the tables before moving on to discuss the results in full.

Tables 5.5 and 5.6 consist of thirteen columns, with the first column indicating the question number. The remaining columns then present the following statistics for each question: mean, mode, median, standard deviation (SD), minimum (Min) and maximum

(Max) scores selected from the Likert scale for that particular question, range, number of responses and finally, the number of teachers who answered 1, 2, 3 or 4. Each row in each table, then, represents one question.

Using this set of data should enable us to determine common areas of agreement among the teachers. For example, answers with a high central tendency (i.e. mean, median and mode score) would suggest that most of the teachers responded to that question positively (e.g. with strongly agree). Likewise, a lower set of scores will indicate that most of the teachers responded negatively (e.g. with strongly disagree). The standard deviation, which is ‘a sort of average of the distance of all the numbers from the mean’ (Brown, 2001: 124); the minimum/maximum scores; and the range show the dispersion of the answers.

In the proceeding discussion on page 67, the results of the teacher survey are examined according to the following four areas: (i) Experience and current duties, (ii) Classroom methodology, (iii) Teacher attitudes and influencing factors, and (iv) EFL tests and English testing in Japan. Throughout the discussion, reference is given to the relevant question numbers on the questionnaire and, in the case of the Likert responses, the mean, mode and standard deviation are also provided in parenthesis.

Table 5.5 Teacher questionnaire: Results from Likert scale questions 4 – 11

Question #	Mean	Mode	Median	SD	Min	Max	Range	N	1	2	3	4
Q4a	3.57	4.00	4.00	0.79	2	4	3	7	0	1	1	5
Q4b	1.57	1.00	1.00	0.79	1	3	3	7	4	2	1	0
Q4c	1.57	2.00	2.00	0.53	1	2	2	7	3	4	0	0
Q4d	3.00	3.00	3.00	0.58	2	4	3	7	0	1	5	1
Q4e	1.71	1.00	2.00	0.76	1	3	3	7	3	3	1	0
Q5	2.43	2.00	2.00	0.53	2	3	2	7	0	4	3	0
Q6	3.71	4.00	4.00	0.49	3	4	2	7	0	0	2	5
Q7	3.86	4.00	4.00	0.38	3	4	2	7	0	0	1	6
Q8a	3.57	4.00	4.00	0.53	3	4	2	7	0	0	3	4
Q8b	2.57	3.00	3.00	0.98	1	4	4	7	1	2	3	1
Q8c	3.29	3.00	3.00	0.76	2	4	3	7	0	1	3	3
Q9a	2.86	3.00	3.00	0.90	1	4	4	7	1	0	5	1
Q9b	3.43	3.00	3.00	0.53	3	4	2	7	0	0	4	3
Q9c	2.29	3.00	2.00	0.76	1	3	3	7	1	3	3	0
Q9d	2.29	3.00	2.00	0.76	1	3	3	7	1	3	3	0
Q9e	3.43	3.00	3.00	0.53	3	4	2	7	0	0	4	3
Q9f	3.00	3.00	3.00	0.58	2	4	3	7	0	1	5	1
Q9g	3.00	3.00	3.00	0.58	2	4	3	7	0	1	5	1
Q9h	3.00	3.00	3.00	0.00	3	3	1	7	0	0	7	0
Q9i	3.57	4.00	4.00	0.53	3	4	2	7	0	0	3	4
Q9j	3.29	3.00	3.00	0.76	2	4	3	7	0	1	3	3
Q9k	2.00	2.00	2.00	0.82	1	3	3	7	2	3	2	0
Q10a	3.86	4.00	4.00	0.38	3	4	2	7	0	0	1	6
Q10b	3.57	4.00	4.00	0.53	3	4	2	7	0	0	3	4
Q10c	3.57	4.00	4.00	0.53	3	4	2	7	0	0	3	4
Q10d	4.00	4.00	4.00	0.00	4	4	1	7	0	0	0	7
Q10e	3.43	3.00	3.00	0.53	3	4	2	7	0	0	4	3
Q10f	3.57	4.00	4.00	0.53	3	4	2	7	0	0	3	4
Q10g	2.86	3.00	3.00	0.69	2	4	3	7	0	2	4	1
Q10h	3.14	3.00	3.00	0.69	2	4	3	7	0	1	4	2
Q11a	3.14	3.00	3.00	0.38	3	4	2	7	0	0	6	1
Q11b	3.00	3.00	3.00	0.58	2	4	3	7	0	1	5	1
Q11c	2.86	3.00	3.00	0.69	2	4	3	7	0	2	4	1
Q11d	3.14	3.00	3.00	0.38	3	4	2	7	0	0	6	1
Q11e	2.86	2.00	3.00	0.90	2	4	3	7	0	3	2	2
Q11f	2.86	3.00	3.00	0.69	2	4	3	7	0	2	4	1
Q11g	2.71	3.00	3.00	0.49	2	3	2	7	0	2	5	0
Q11h	3.57	4.00	4.00	0.53	3	4	2	7	0	0	3	4
Q11i	3.14	3.00	3.00	0.69	2	4	3	7	0	1	4	2
Q11j	2.71	3.00	3.00	0.49	2	3	2	7	0	2	5	0
Q11k	3.14	3.00	3.00	0.69	2	4	3	7	0	1	4	2
Q11l	2.57	2.00	2.00	0.79	2	4	3	7	0	4	2	1

Table 5.6 Teacher questionnaire: Results from Likert scale questions 12 – 14

Question #	Mean	Mode	Median	SD	Min	Max	Range	N	1	2	3	4
Q12a	3.43	4.00	4.00	0.79	2	4	3	7	0	1	2	4
Q12b	3.86	4.00	4.00	0.38	3	4	2	7	0	0	1	6
Q12c	2.71	3.00	3.00	0.76	2	4	3	7	0	3	3	1
Q12d	3.14	3.00	3.00	0.69	2	4	3	7	0	1	4	2
Q12e	2.71	3.00	3.00	0.49	2	3	2	7	0	2	5	0
Q12f	3.29	3.00	3.00	0.76	2	4	3	7	0	1	3	3
Q12g	3.00	3.00	3.00	0.82	2	4	3	7	0	2	3	2
Q12h	2.86	3.00	3.00	0.69	2	4	3	7	0	2	4	1
Q12i	2.71	3.00	3.00	0.49	2	3	2	7	0	2	5	0
Q12j	2.71	2.00	3.00	0.76	2	4	3	7	0	3	3	1
Q12k	2.86	2.00	3.00	0.90	2	4	3	7	0	3	2	2
Q12l	3.14	3.00	3.00	0.69	2	4	3	7	0	1	4	2
Q12m	3.14	3.00	3.00	0.69	2	4	3	7	0	1	4	2
Q12n	2.86	3.00	3.00	0.69	2	4	3	7	0	2	4	1
Q12o	3.29	3.00	3.00	0.76	2	4	3	7	0	1	3	3
Q12p	3.43	3.00	3.00	0.53	3	4	2	7	0	0	4	3
Q12q	3.43	3.00	3.00	0.53	3	4	2	7	0	0	4	3
Q13a (S)	1.40	1.00	1.00	0.55	1	2	2	5	3	2	0	0
Q13b (S)	2.17	2.00	2.00	0.75	1	3	3	6	1	3	2	0
Q13c (S)	2.67	3.00	3.00	0.82	1	3	3	6	1	0	5	0
Q13d (S)	1.83	2.00	2.00	0.41	1	2	2	6	1	5	0	0
Q13e (S)	1.83	2.00	2.00	0.41	1	2	2	6	1	5	0	0
Q13f (S)	1.83	2.00	2.00	0.41	1	2	2	6	1	5	0	0
Q13g (S)	3.17	4.00	3.50	1.17	1	4	4	6	1	0	2	3
Q13a (W)	3.33	4.00	3.50	0.82	2	4	3	6	0	1	2	3
Q13b (W)	2.43	2.00	2.00	0.98	1	4	4	7	1	3	2	1
Q13c (W)	2.33	2.00	2.00	1.03	1	4	4	6	1	3	1	1
Q13d (W)	3.43	3.00	3.00	0.53	3	4	2	7	0	0	4	3
Q13e (W)	3.29	3.00	3.00	0.49	3	4	2	7	0	0	5	2
Q13f (W)	3.29	3.00	3.00	0.49	3	4	2	7	0	0	5	2
Q13g (W)	3.00	3.00	3.00	1.10	1	4	4	6	1	0	3	2
Q14a	2.86	3.00	3.00	0.69	2	4	3	7	0	2	4	1
Q14b	3.43	4.00	4.00	0.79	2	4	3	7	0	1	2	4
Q14c	2.50	2.00	2.50	0.55	2	3	2	6	0	3	3	0
Q14d	3.43	3.00	3.00	0.53	3	4	2	7	0	0	4	3
Q14e	2.29	2.00	2.00	0.49	2	3	2	7	0	5	2	0
Q14f	2.71	2.00	3.00	0.76	2	4	3	7	0	3	3	1
Q14g	2.57	1.00	3.00	1.27	1	4	4	7	2	1	2	2
Q14h	3.00	3.00	3.00	0.58	2	4	3	7	0	1	5	1
Q14i	2.67	3.00	3.00	1.03	1	4	4	6	1	1	3	1
Q14j	2.43	3.00	3.00	0.98	1	3	3	7	2	0	5	0

(i) Experience and current duties

Questions 1 and 2 on the teacher questionnaire simply recorded relevant biodata – i.e. how long the teachers had taught English and which courses they were now teaching. Of the seven teachers in the sample, five had 10 or fewer years experience and two had over 25 years experience. A total of three teachers were teaching on the general English course offered by their respective schools ('English I/II'); one was teaching a grammar course in addition to the general course; one was teaching reading in addition to the general course and the remaining two teachers were teaching on all three courses. The three youngest teachers in the sample also had responsibility for Oral Communication classes.

(ii) Classroom methodology

Within this category, the teachers' responses to items 3, 4, 11 and 12 on the questionnaire have been grouped together for analysis. Specifically, these questions refer to the preferred teaching practices of the survey respondents, and also provide insight into the type of language resources and materials teachers consider being most appropriate for Japanese learners.

When asked to state which language they generally used when teaching English (Q3), six out of the seven teachers reported that they used mainly Japanese as the language of instruction, with one saying that he used half English and half Japanese. However, one teacher also commented that it depended on the ability of the group, stating that with higher-level groups she tried to deliver most of the lesson in the target language, if possible. The choice of L1 as the language of instruction is also consistent with the way in which classes are generally organised, with most teachers stating that they preferred to adopt a teacher-led approach (Q4a/mean=3.57/mode=4.00/SD=0.79), whilst further admitting that they rarely implemented pair work activities. Indeed, four

respondents reported that they never implemented pair work activities (Q4b/mean=1.57/mode=1.00/SD=0.79).

In terms of materials and resources, the majority of teachers thought that prepared worksheets were most suitable for Japanese high school students (Q11h/mean=3.57/mode=4.00/SD=0.53) and they were also generally in favour of using past examination papers (Q11g/mean=2.71/mode=3.00/SD=0.49).

Most teachers also agreed that reading aloud was an effective language learning strategy, either with the teacher assuming the role of reader (Q12a/mean=3.43/mode=4.00/SD=0.79), or with the learner assuming the role of reader (Q12b/mean=3.86/mode=4.00/SD=0.38). Multiple-choice grammar/vocabulary exercises and reading comprehension exercises were also considered to be important (Q12l & Q12m/mean=3.14/mode=3.00/SD=0.69), as were translation exercises – both L2 to L1 (Q12o/mean=3.29/mode=3.00/SD=0.76) and L1 to L2 (Q12p/mean=3.43/mode=3.00/SD=0.53).

It could be suggested, then, that the results concerning teaching methods and use of materials presented so far, serve to confirm the findings of the observation sessions. In other words, to a large extent teachers would appear to be doing at classroom level what they in fact claim to be doing. However, evidence to the contrary also exists as several contradictions occur throughout the questionnaire. For example, whilst there was strong agreement that speaking activities in pairs or groups should be used for teaching English to Japanese high school students (Q12q/mean=3.43/mode=3.00/SD=0.53), this did not correspond with what teachers actually reported doing. As already noted (and indeed, observed), teacher-centred classrooms are more common in the Japanese context.

It should also be carefully noted that, with regards to teaching resources (Q11), all but two of the options (adapted video and English word games/puzzles) returned mode scores of at least 3.00. Thus, contained within the list of language resources teachers believed should be used were listening exercises using authentic speech (Q11a/mean=3.14/mode=3.00/SD=0.38), authentic video (Q11d/mean=3.14/mode=3.00/SD=0.38) and language learning websites/e-mail (Q11i/mean=3.14/mode=3.00/SD=0.69). Again this tends to conflict somewhat with the findings of the observation sessions. However, as noted earlier in the discussion when we looked at the results of the observation sessions, the fact that teachers prefer to adopt a teacher-led, grammar-based approach may not necessarily be an indication of negative washback from the examinations.

(iii) Teacher attitudes and influencing factors

Items 7, 8, 9, 10 and 14 on the questionnaire can all be seen to relate to teacher attitudes or to factors (such as external tests) which may serve to inform and shape those attitudes. In view of the preferred teaching methodology in Japan, this area also produced a rather contradictory set of results. For example, most teachers believed that it was very important for their learners to develop communicative ability (Q7/mean=3.86/mode=4.00/SD=0.38) but that native speaker English teachers should teach listening and speaking skills (Q8a/mean=3.57/mode=4.00/SD=0.53). Being able to communicate with English-speaking visitors who come to Japan (Q9i/mean=3.57/mode=4.00/SD=0.53) was also seen as one of the most important reasons for students to study English – indeed, this was regarded as being slightly more important than studying in order to pass university entrance exams (Q9b/mean=3.43/mode=3.00/SD=0.53). As a result, when teachers were asked to state the areas which they thought were most important for Japanese learners of English, the highest central tendencies were reported within the categories of speaking (Q10d/mean=4.00/mode=4.00/SD=0.00) and listening

comprehension (Q10a/mean=3.86/mode=4.00/SD=0.38). It should be noted however, that no single area really stood out as being more important than any other in this question with mode scores of 4.00 also recorded for writing, reading comprehension and vocabulary. Grammar (Q10e; mean=3.43) was still considered to be an important part of English study but interestingly, fewer teachers (n=3) strongly agreed that it should be seen as one of the most important aspects of language study. Given the nature of these responses, then, it would seem to suggest that whilst teachers recognise the importance of developing their learners' communicative skills, they are perhaps less willing in practice to use techniques which might aid that development. Inconsistencies again emerge with regards to the practice and development of speaking skills when we consider that all teachers in the sample strongly agreed that speaking was one of the most important areas for learners, yet, for the most part, continue to conduct lessons almost entirely in the first language.

Teaching experience (Q14b/mean=3.43/mode=4.00/SD=0.79) and past language learning experiences (Q14d/mean=3.43/mode=3.00/SD=0.53) were regarded as the being the most influential factors behind current teaching approaches. Interestingly, external examinations were reported as having less influence, although, despite the low mode score of 1.00, answers were fairly evenly distributed in this case (Q14g/mean=2.57/mode=1.00/SD=1.27).

(iv) EFL tests and English testing in Japan

This final area again looks at teacher attitudes and opinions, but this time more specifically in relation to the English tests taken by Japanese high school students. Related items on the questionnaire are numbers 5, 6, 13, 15 and 16. Question 5 asked teachers if they thought it necessary for students to attend a cram school in order to pass university examinations, as is common in Japan. Responses were more or less divided,

with four teachers feeling that it was not so important and three believing that it was important (Q5/mean=2.43/mode=2.00/SD=0.53).

With regards to the nature of language tests, most teachers believed it was very important to test writing and speaking (Q6/mean=3.71/mode=4.00/SD=0.49) but also agreed that the regional tests (Q13a(S)/mean=1.40/mode=1.00/SD=0.55) and university entrance examinations (Q13d, e and f(S) with identical central tendencies of mean=1.83 /mode=2.00/SD=0.41) – which arguably represent the most important areas of assessment – did not provide an accurate assessment of speaking ability. On the other hand, the TOEIC (Q13g(S)/mean=3.17/mode=4.00/SD=1.17) was thought to provide a more accurate assessment of speaking ability, with just one teacher disagreeing. The high central tendency recorded here perhaps indicates a lack of familiarity with this particular test, however, as the TOEIC does not actually contain a speaking component.

The final two questions on the survey were open-ended and asked for, (i) teachers' views on the introduction of a listening component to the Center Test which will come into effect in 2006 (question 15), and (ii) further comments on the current range of tests taken by their learners (question 16). Unfortunately, only three teachers chose to answer these questions – the remaining four declining to comment. With regard to question 15, one teacher felt unable to comment due to the fact that none of her students were actually preparing for the Center Test next year. The other two teachers reported that the addition of a listening section to the Center Test would not lead them to change the way they prepared students for the exam. One commented, "I will teach the students in the same way as usual" whilst the other believed that the most effective way to prepare students for the Center Test was to encourage them to "learn many words and sentences by heart".

Question 16 provided two contrasting views over the effects of current tests. The more experienced teachers believed that these tests have a positive effect on learning as well as providing an accurate measure of English ability. The importance attached to the university entrance examinations was also noted – indeed these two respondents felt that they perhaps represented the main reason why most students actually study English. The younger teacher who replied to this question, however, disagreed somewhat. She felt that without changes to the Center Test or university examinations, it would be difficult to change (and improve) teaching habits. She made the observation that in the current climate even if she wanted to teach something “more important” she was unable to do so, commenting, “I know there is a lot of grammar on the test and I know what kind of grammar will be on the test, so I feel I have to teach that to the students because that is what they expect to learn.” This clearly implies that in some cases at least, external tests may have a negative impact on teaching and learning.

5.2.2 Student questionnaire #1

Let us now consider the first student questionnaire (see Appendix VIII) using a similar method of analysis. To recapitulate, this questionnaire was completed by group 4 (n=26)¹ and group 5 (n=20). The main objectives were to measure students’ attitudes to second language study and to determine the extent of examination washback on learning. Likert scale items again made up the bulk of the questionnaire. The results are presented in full in tables 5.7 and 5.8 (pp.73-74), using the same set of descriptive statistics that were adopted for the teacher questionnaire. In view of the larger number of respondents who made up the student sample, however, additional frequency and percentage figures are presented, where appropriate, to enable us to see more clearly the distribution of answers. It should also be noted that the data collected from both groups are examined together in order that the two sets of responses can be compared more easily.

¹ One student was absent from group 4 during the administration of the first questionnaire.

The data collected from the student questionnaire have been broken down into the following three areas of analysis: (i) Learning experience, (ii) Attitudes towards English, and (iii) Preferred learning styles. Once again, reference is made to the relevant numbers on the questionnaire and central tendency statistics are provided in parenthesis.

Table 5.7 Student questionnaire #1: Results from Likert scale questions for group 4 (n=26)

Question #	Mean	Mode	Median	SD	Min	Max	Range	N	1	2	3	4
4a	2.62	3.00	3.00	0.85	1	4	4	26	3	7	13	3
4b	2.85	3.00	3.00	0.97	1	4	4	26	3	5	11	7
4c	2.73	3.00	3.00	0.87	1	4	4	26	3	5	14	4
4d	1.92	1.00	2.00	0.93	1	4	4	26	10	10	4	2
4e	2.46	3.00	2.50	1.07	1	4	4	26	6	7	8	5
4f	2.00	1.00	2.00	0.89	1	4	4	26	9	9	7	1
4g	2.04	1.00	2.00	1.08	1	4	4	26	12	3	9	2
4h	2.27	2.00	2.00	0.83	1	4	4	26	5	10	10	1
4i	2.69	3.00	3.00	0.93	1	4	4	26	3	7	11	5
4j	2.38	2.00	2.00	0.94	1	4	4	26	5	9	9	3
5a	3.35	3.00	3.00	0.75	1	4	4	26	1	1	12	12
5b	3.00	3.00	3.00	0.69	2	4	3	26	0	6	14	6
5c	3.15	3.00	3.00	0.78	1	4	4	26	1	3	13	9
5d	3.42	4.00	4.00	0.76	1	4	4	26	1	1	10	14
5e	3.12	3.00	3.00	0.59	2	4	3	26	0	3	17	6
5f	3.12	3.00	3.00	0.59	2	4	3	26	0	3	17	6
5g	3.00	3.00	3.00	0.94	1	4	4	26	2	5	10	9
5h	3.15	3.00	3.00	0.78	1	4	4	26	1	3	13	9
6a	3.12	3.00	3.00	0.59	2	4	3	26	0	3	17	6
6b	2.65	3.00	3.00	0.75	1	4	4	26	1	10	12	3
6c	2.96	3.00	3.00	0.82	1	4	4	26	2	3	15	6
6d	2.73	3.00	3.00	0.83	1	4	4	26	2	7	13	4
6e	2.73	3.00	3.00	0.72	1	4	4	26	1	8	14	3
6f	3.23	3.00	3.00	0.82	1	4	4	26	1	3	11	11
6g	2.85	3.00	3.00	0.73	1	4	4	26	1	6	15	4
6h	2.38	2.00	2.00	0.64	1	4	4	26	1	15	9	1
6i	2.88	3.00	3.00	0.71	1	4	4	26	1	15	16	4
6j	3.31	3.00	3.00	0.68	2	4	3	26	0	3	12	11
6k	3.31	3.00	3.00	0.68	2	4	3	26	0	3	12	11
6l	2.85	3.00	3.00	0.61	2	4	3	26	0	7	16	3
6m	2.81	3.00	3.00	0.63	2	4	3	26	0	8	15	3
6n	2.73	3.00	3.00	0.67	2	4	3	26	0	10	13	5
6o	2.88	3.00	3.00	0.77	1	4	4	26	1	6	14	5
7a	3.31	3.00	3.00	0.74	1	4	4	26	1	1	13	11
7b	2.69	3.00	3.00	0.55	2	4	3	26	0	9	16	1
7c	2.65	3.00	3.00	0.63	2	4	3	26	0	11	13	2
7d	2.69	3.00	3.00	0.79	1	4	4	26	2	7	14	3

Table 5.8 Student questionnaire #1: Results from Likert scale questions for group 5 (n=20)

Question #	Mean	Mode	Median	SD	Min	Max	Range	N	1	2	3	4
4a	3.00	3.00	3.00	0.56	2	4	3	20	0	3	14	3
4b	2.40	3.00	2.50	0.94	1	4	4	20	4	6	8	2
4c	2.10	2.00	2.00	0.91	1	4	4	20	5	10	3	2
4d	1.74	1.00	2.00	0.73	1	3	3	19	8	8	3	0
4e	2.45	2.00	2.00	1.00	1	4	4	20	3	9	4	4
4f	1.65	1.00	1.50	0.75	1	3	3	20	10	7	3	0
4g	1.50	1.00	1.00	0.61	1	3	3	20	11	8	1	0
4h	1.70	1.00	1.50	0.80	1	3	3	20	10	6	4	0
4i	2.60	3.00	3.00	0.88	1	4	4	20	3	4	11	2
4j	2.25	2.00	2.00	0.97	1	4	4	20	5	7	6	2
5a	3.30	4.00	3.50	0.86	1	4	4	20	1	2	7	10
5b	2.75	4.00	3.00	1.07	1	4	4	20	3	5	6	6
5c	2.85	3.00	3.00	1.04	1	4	4	20	3	3	8	6
5d	3.30	4.00	4.00	0.92	1	4	4	20	1	3	5	11
5e	2.70	2.00	3.00	0.98	1	4	4	20	2	7	6	5
5f	2.80	3.00	3.00	0.83	1	4	4	20	1	6	9	4
5g	2.75	3.00	3.00	0.97	1	4	4	20	3	3	10	4
5h	2.70	3.00	3.00	0.98	1	4	4	20	3	4	9	4
6a	2.95	3.00	3.00	0.76	1	4	4	20	1	3	12	4
6b	2.45	2.00	2.00	0.94	1	4	4	20	3	8	6	3
6c	2.90	3.00	3.00	0.79	2	4	3	20	0	7	8	5
6d	2.79	3.00	3.00	0.63	2	4	3	19	0	6	11	2
6e	2.79	4.00	3.00	1.08	1	4	4	19	2	7	3	7
6f	2.74	2.00	3.00	0.93	1	4	4	19	1	8	5	5
6g	2.95	3.00	3.00	0.71	2	4	3	19	0	5	10	4
6h	2.53	3.00	3.00	0.70	1	4	4	19	1	8	9	1
6i	2.84	3.00	3.00	0.83	1	4	4	19	1	5	9	4
6j	3.21	4.00	3.00	0.92	1	4	4	19	1	3	6	9
6k	3.21	4.00	3.00	0.92	1	4	4	19	1	3	6	9
6l	2.68	3.00	3.00	0.67	2	4	3	19	0	8	9	2
6m	2.63	3.00	3.00	0.76	1	4	4	19	1	7	9	2
6n	3.21	3.00	3.00	0.71	2	4	3	19	0	3	9	7
6o	2.79	3.00	3.00	0.63	2	4	3	19	0	6	11	2
7a	3.37	4.00	3.00	0.68	2	4	3	19	0	2	8	9
7b	2.74	3.00	3.00	0.87	1	4	4	19	1	7	7	4
7c	2.42	3.00	3.00	0.84	1	4	4	19	3	6	9	1
7d	2.47	2.00	2.00	0.84	1	4	4	19	2	8	7	2

(i) Learning experience

In group 4, most of the learners (73%, n=19) had studied English since the first grade of junior high school (aged 12 – 13 years old), whereas the remainder (27%, n=7) reported having had some earlier experience at kindergarten, elementary school or at home. Likewise, the majority of the learners in group 5 (75%, n=15) also said they had first started to study English during the first grade of junior high school. Only 2 learners from each group reported having extra classes outside of school.

(ii) Attitudes towards English

Responses to items 1, 4 and 5 on the questionnaire have been included for analysis in this section. Question 1 asked students to place the subjects they studied at school in order of importance, with '1' indicating the most important, '10' the least important. In terms of the subjects that are of most concern to us here, i.e. English IR/IIR (reading), English IG/IIG (grammar) and OC I/II (Oral Communication), 12 students (46%) from group 4 considered English grammar classes to be the most important of their three language subjects and 11 students (42%) thought OC was the most important. Only 3 students (12%) in this group placed reading above OC or grammar. Group 5, on the other hand, considered Oral Communication to be far less significant with just 5 students (25%) ranking it higher than English reading and grammar. However, nobody ranked it higher than sixth overall in terms of importance. Of course, the data collected from question 1 needs to be treated with some caution as, assuming students answered truthfully, the results may do little more than illustrate that the motivation of one group (i.e. group 4) is greater than the other (i.e. group 5) when it comes to second language study.

Question 4 examined the possible motivating factors behind the learners' study of English. External examinations, perhaps not surprisingly, had the greater influence on group 4, with 18 students (69%) agreeing that both the university entrance tests and the

monthly regional tests constituted the main reasons for studying English (Q4b/mean=2.85/mode=3.00/SD=0.97; Q4c/mean=2.73/mode=3.00/SD=0.87). For group 5, most students (85%, n=17) reported that they were studying English because they “have to, in order to graduate” (Q4a/mean=3.00/mode=3.00/SD=0.56). University examinations were seen as a motivating influence by some of the learners in group 5, however, although responses were evenly distributed between agreement and disagreement. This can be explained by the fact that some learners will indeed sit entrance examinations, although they are most likely to be for admission to two-year courses at junior college level

There is some evidence to suggest that there may be other motivating factors, besides tests, when it comes to second language study. For example, the majority of students in group 4 (62%, n=16) said that they wanted to study English in order to communicate with English-speaking visitors to Japan (Q4i/mean=2.69/mode=3.00/SD=0.93), with similar results (65%, n=13) also reported in group 5 (Q4i/mean=2.60/mode=3.00/SD=0.88).

As with the corresponding responses on the teacher questionnaire, learners also considered listening comprehension and speaking to be the most important elements of language learning, both returning the two highest central tendencies for each group (Q5a & Q5d). Of most relevance amongst the data collected from question 5, however, is the fact that the statistics show that group 4 attached more importance than group 5 to the remaining areas of writing, reading comprehension, grammar, vocabulary, spelling and pronunciation/stress – all key areas of assessment on the regional tests and university entrance examinations. This is particularly significant bearing in mind that these represent examples of “high-stakes” tests for this group of learners, whereas the same cannot be said for group 5.

(iii) Preferred learning styles

As already noted above, external examinations appear to exert greater influence on the learners in group 4, particularly in terms of what they consider to be the most important areas of study. Let us now examine whether or not there is a significant difference in the preferred learning styles of group 4 (for whom external examinations hold greater importance) and group 5.

In view of what has already been said about the two groups and their targeted examinations, it was predicted that both sets of learners would indicate a strong preference for multiple-choice reading questions and multiple-choice/grammar/vocabulary exercises as test preparation activities. Group 4, on the other hand, would be more likely to use past exam papers and practice translation exercises.

The results actually indicate that group 4 learners reported that they found multiple-choice activities slightly more helpful for exam preparation than group 5 (see Q6l & Q6m). Although mode scores of 3.00 were recorded in each category for both groups, the responses given by group 5 are again more evenly distributed indicating that opinions were more divided.

Past exam papers (see Q6f) were indeed more popular with group 4, with most of the learners (85%, n=22) agreeing that they were helpful in preparing for exams (Q6f/mean=3.23/mode=3.00/SD=0.82). In group 5, roughly half of the learners (53%, n=10) who responded to the same question agreed (Q6f/mean=2.74/mode=2.00/SD=0.93). More learners in group 4 (77%, n=20) also reported using past exam papers during their own private study time than in group 5 (35%, n=7).

There was further evidence of possible test washback in the case of group 4 with the majority of the learners (88%; n=23) strongly indicating that they thought learning through translation was the most effective way in which to prepare for examinations (Q6j & 6k/mean=3.31/mode=3.00/SD=0.68). Once again though, any such conclusion must be reached somewhat tentatively, particularly in view of the fact that group 5 also regarded translation as being the most useful method of test preparation. Whilst most students said they would rather prepare for tests during class time with the help of a teacher, translation activities were also used during private study by 63% of the learners in group 4 and 45% of the learners in group 5.

5.2.3 Student questionnaire #2

The final student questionnaire (see Appendix IX) was administered to group 4 the week after they had taken the paired speaking test. In consideration of the second research question, the aim was to discover whether the test had had the desired effect on learning – i.e. was it indeed possible to nurture positive washback by changing assessment techniques?

This time the questionnaire consisted entirely of Likert scale items – each one a single statement which the learner was asked to rate along the previously adopted scale of 1 (strongly disagree) to 4 (strongly agree). There were nineteen statements in all; with additional space for any further comments about the test that the learner wanted to add. The results of the questionnaire are again illustrated through the use of descriptive statistics and are presented in table 5.9 below along with each of the nineteen statements.

Table 5.9 Results of student questionnaire #2 (Group 4, n=27)

Question #	Mean	Mode	Median	SD	Min	Max	Range	N	1	2	3	4
1	2.67	3.00	3.00	0.68	1	4	4	27	1	9	15	2
2	2.70	3.00	3.00	0.67	1	4	4	27	1	8	16	2
3	2.70	3.00	3.00	0.89	1	4	4	27	4	9	11	3
4	2.93	3.00	3.00	0.87	1	4	4	27	1	8	10	8
5	1.81	1.00	2.00	0.83	1	4	4	27	11	11	4	1
6	2.67	3.00	3.00	0.92	1	4	4	27	3	8	11	5
7	2.74	3.00	3.00	0.66	1	4	4	27	1	7	17	2
8	3.11	3.00	3.00	0.80	1	4	4	27	1	4	13	9
9	2.67	3.00	3.00	0.83	1	4	4	27	2	9	12	4
10	2.41	2.00	2.00	0.80	1	4	4	27	2	15	7	3
11	2.19	2.00	2.00	0.74	1	4	4	27	3	18	4	2
12	2.37	2.00	2.00	0.63	1	3	3	27	2	13	12	0
13	2.59	2.00	3.00	0.93	1	4	4	27	3	10	9	5
14	3.11	3.00	3.00	0.80	2	4	3	27	0	7	10	10
15	2.63	3.00	3.00	0.74	1	4	4	27	1	11	12	3
16	3.22	3.00	3.00	0.70	1	4	4	27	1	1	16	9
17	2.33	2.00	2.00	0.73	1	4	4	27	3	13	10	1
18	3.48	4.00	4.00	0.58	2	4	3	27	0	1	12	14
19	1.89	2.00	2.00	0.80	1	4	4	27	9	13	4	1

- (1) I enjoyed doing the speaking test with a partner.
- (2) Studying for the speaking test improved my English.
- (3) I made a greater effort to speak English in the weeks leading up to the test.
- (4) It was important to do well on test.
- (5) It was possible to do well on the test without much preparation.
- (6) If the test was included as part of the Seibu Moshi or Center Test I would do more preparation.
- (7) I enjoyed practising for the test during class.
- (8) It was important to practise for the test during class.
- (9) There wasn't enough time to practise for the test.
- (10) It is more important to practise other language skills during class.
- (11) When not in class, it is better to study for a speaking test on your own.
- (12) When not in class, it is better to study for a speaking test with friends.
- (13) I would have liked to have had access to more materials for practice out of class.
- (14) It is difficult to study for a speaking test in your own time without a teacher.
- (15) It wasn't necessary to study for the speaking test in my own time.
- (16) In order to do well on the test it is necessary to memorise key phrases/vocabulary.
- (17) It is only necessary to speak English during Oral Communication classes.
- (18) I want to improve my English speaking skills.
- (19) It isn't important for me to speak English, so speaking shouldn't be tested.

The above responses, together with the informal feedback received from learners following the test, suggest that in general the paired test format was favourably received. One learner commented, “It is important to speak English with each other, so it was much better to do the test in pairs” whilst another added, “The speaking test was useful to practise English conversation. It was good fun to do the speaking test with a partner – it is important to practise English even if I don’t have the opportunity to travel much in the future”. These sentiments appear to be supported by the responses given to statements 1 and 7, as shown above in table 5.9. However, one learner disagreed, saying that whilst she thought it was a very good test, the one-to-one format (i.e. interlocutor/assessor – examinee) afforded a better opportunity to practise for the STEP speaking test.

There is further evidence to suggest that the test had a positive washback effect on learners and the learning process. For example, the majority of learners (67%, n=18) agreed that doing the test had improved their English (Q2) and, encouragingly, a reasonably high proportion of the group (52%, n=14) also responded positively to the third statement: I made a greater effort to speak English in the weeks leading up to the test.

As noted previously, tests often provide a source of extrinsic motivation for the learners in the sample (and indeed for Japanese learners in general). This also proved to be the case with the new speaking test, as most students (67%, n=18) agreed that ‘it was important to do well in the test’ (Q4) and, in turn, that it was difficult to do well without much preparation (Q5). Having had the opportunity to observe the learners during Oral Communication classes in the weeks prior to the test, I would say that they appeared more motivated, particularly during paired speaking activities – an observation supported by the responses recorded for number 8, which indicated that 22 learners (81%) agreed that ‘it was important to practise for the test during class’. Practising during class time

with the support of a teacher also appeared to be the preferred mode of preparation for the speaking test (see, for example, questions 10, 11, 12 and 14).

One possible instance of negative washback is perhaps related more to the actual design of the test. For example, the highest recorded mean score of 3.22 was recorded for number 16, which read: In order to do well on the test it is necessary to memorise key phrases/vocabulary. Whilst this would perhaps not really be classed as a particularly desirable effect of any speaking test, it should be remembered that having carefully considered the overall level of the students in order to give them tasks that they would realistically be able to perform, this situation was to a large extent unavoidable. It is possible that with higher level learners, faced with a range of more challenging tasks, memorisation of key phrases would be less of an option.

5.3. Conclusion and implications

At various points throughout this study, reference has been made to the mismatch that occurs between the stages of curriculum planning and classroom implementation. In other words, the ‘official’ communicative syllabus aims of the Japanese Ministry of Education (MEXT) remain at odds with the more grammar-based methodology adopted by teachers at classroom level. It has also been claimed that English teachers in Japan find themselves in an unenviable position in which the constraints imposed by the examination-driven “hidden syllabus” prevent them from implementing, in practice, communicative methodology. The findings of this study, however, indicate that the existence of a direct causal relationship between language testing methodology and the preferred grammar-based approach of teachers may not be that straightforward.

It would appear that whilst current examinations may help to determine what is actually taught, they may have rather less influence over how teachers teach. Thus, in

terms of teaching, the evidence presented here tends to support the earlier findings of Wall & Alderson (1993) and Cheng (1997), notably that the washback effect can be observed at the level of content but not at the level of methodology. This is demonstrated by the fact that the materials used by the teachers in this study consisted of grammar and translation exercises of the type often found on the majority of external tests taken by the learners in the sample. When teaching groups that did not have to sit these tests, however, both content and methodology remained consistent with the grammar-based approach.

The question remains then, whether changes to examinations alone would have the desired washback effect on teaching methodology. Certainly, it would appear that in the case of some teachers a more communicative approach to language testing would encourage a similarly communicative approach to language teaching. In other instances, however, teachers seem less likely to embrace new ideas or methods simply because the format of an examination changes, preferring instead to continue using the grammar-based methodology with which they are most familiar. Even in this relatively small scale study attitudes varied, with one teacher in particular suggesting that if changes were made to tests, she would be in a better position to change not only the content of her lessons, but also her approach to teaching (see discussion on pages 71-72).

The paired speaking test that was trialled received some favorable feedback from the teachers in the sample group. For example, teachers A and B thought that it provided an excellent means of assessment in the area of speaking, and were particularly impressed by the paired format – teacher A commenting that “it was good that the person scoring the test didn’t have to concentrate on directing the learner’s conversation too much”. As a result, both teachers felt confident that they would be able to score the test (and thus assume the role of assessor). However, they remained doubtful about whether the test would cause them to change the format of their lessons if it was included as a component

of one of the existing external examinations. As such it would appear that, unfortunately, the paired speaking test introduced here would not have the intended washback effect on teaching methodology in this particular sample context.

It thus becomes apparent that the potential for nurturing positive washback by exacting changes to examinations certainly exists – indeed, in the Japanese context any innovations within the area of assessment must surely be encouraged. However, changes should perhaps occur in combination with comprehensive in-service teacher training programmes designed to increase familiarity with communicative teaching and testing methodology. If MEXT are serious about promoting communicative pedagogy at classroom level, initial teacher training courses must also accord with that philosophy. Collins (1999) notes that successful implementation of the communicative approach requires teachers to have ‘a very high degree of competence in all four skills of speaking, listening, reading and writing in English’. However, she goes on to observe that,

...many teachers in Japanese high schools lack the phonological and sociolinguistic knowledge which would give them the confidence and expertise required to deal with the complexities of intonation and sentence rhythm. The curriculum they are offered as English majors in national universities of Education, itself dictated by the Monbusho (MEXT) is not sufficiently specialised or focused, and does not prepare teachers adequately in these areas to be able to deal with the functionally based syllabus required by the communicative approach.

(Collins 1999: retrieved from the World Wide Web)

Although grammar-based instruction remains the dominant methodology, it should be noted that the teachers and learners in the sample both recognised the importance of developing listening and speaking skills – indeed, evidence that the paired speaking test had had a positive washback effect on learning within these two areas was observed earlier. However, as implied in the washback hypotheses put forward by Alderson &

Wall (1993: 120-121), if a test influences what teachers teach and how teachers teach, those two variables alone will clearly relate to the amount of influence a test has on learning. In other words, teachers, rather than learners, are far more likely to be able to effect the necessary changes in order to improve the actual learning process, at least within the context of the classroom. For the most part, the notion of learner autonomy, in which ‘students should have a large amount to say about what, how, and how fast they learn’ (Bailey 1999: 41) remains incompatible with the traditional educational philosophy in Japan.

In summary, then, the results of this study suggest that it is possible to improve learning by employing direct testing techniques. However, the official syllabus aims of MEXT will only be fully realised when positive washback can be observed at the level of both teaching and learning. In order to examine more fully whether beneficial washback can be achieved within the area of teaching, it is suggested that future research should be conducted using a larger sample group and that data should be collected longitudinally, where possible. Both of these considerations were beyond the practicalities of this particular study. However, unless more direct testing techniques are employed on a larger number of “high stakes” examinations in Japan, and the effects that such changes have on teachers and teaching can be observed over a much longer period of time, it will be difficult to observe the true extent of washback in this area.

Appendix I

National Center for University Entrance Examinations (NCUEE) 'Center Test' 2004 sample questions

英 語

(解答番号 ~)

第1問 次の問い(A・B)に答えよ。(配点 16)

A 次の問い(問1・問2)において、下線部(a)・(b)の単語のアクセント(強勢)の位置が正しい組合せを、それぞれ下の①~④のうちから一つずつ選べ。

問1

The huge (a)network of canals is a (b)fascinating feature of the old city.

- ① (a) nétwork (b) fascináting ② (a) netwórk (b) fáscinating
③ (a) netwórk (b) fascináting ④ (a) nétwork (b) fáscinating

問2

The president's (a)comments on energy resources were general; she didn't deal with (b)particular problems.

- ① (a) comménts (b) párticular ② (a) cómments (b) partícular
③ (a) comménts (b) partícular ④ (a) cómments (b) párticular

第2問 次の問い(A～C)に答えよ。(配点 38)

A 次の問い(問1～10)の ～ に入れるのに最も適当なものを、それぞれ下の①～④のうちから一つずつ選べ。

問1 Does this answer sense to you? I can't understand it.

- ① give ② cause ③ mean ④ make

問2 Taro is now devoting all his time and energy English.

- ① studying ② to studying
③ to study ④ study

問3 Every winter, colds are at schools.

- ① familiar ② popular ③ broad ④ common

問4 Akiko had her bag and lost all her money.

- ① steal ② stole ③ stealing ④ stolen

問5 By the time the 2002 World Cup was held, soccer already become a leading sport in Japan.

- ① had ② has ③ was ④ is

B 次の問い(問1～3)の会話の ～ に入れるのに最も適当なものを、それぞれ下の①～④のうちから一つずつ選べ。

問1 Osamu: It's very hot and humid today. Will it be any better tomorrow?

Betty: I heard that it's going to be even worse!

Osamu:

- ① I'm sorry that it's going to be cooler.
- ② That's too bad. I like humid weather.
- ③ It's a shame that it's so dry.
- ④ Oh, no. I'm planning to go hiking.

問2 Jack: I'm hungry. Shall we go for lunch now?

Ken: There isn't time. The meeting is about to start.

Jack:

Ken: That's right. Let's hurry. Everyone must be waiting.

- ① Why will we be late for the meeting?
- ② When did it start?
- ③ What? No lunch today?
- ④ Where would you like to eat?

問3 Janet: When do we have to hand in the report?

Ichiro: I don't remember.

Janet: Oh, good. Could you let me know what you find out from her?

- ① I'll go and ask the teacher.
- ② Did the teacher already tell us?
- ③ I don't know when it's due.
- ④ I've just started it.

第3問 次の問い(A～C)に答えよ。(配点 34)

- A 次の文章の ・ に入れるのに最も適当な語句を、それぞれ下の①～④のうちから一つずつ選べ。

When English-speaking people talk about “hot” food, are they saying the food is spicy like curry, or are they talking about its temperature, as in “hot” coffee? These two different meanings of “hot” may seem confusing to Japanese students, but , the word is the right one for describing the way the body responds to spice and heat. A simple explanation would go something like this: when we eat or drink, the same nerves in the mouth react both to spicy chemicals in the food and to a rise in temperature. The English expression, , reflects this fact about the human body.

- ① in contrast to this
- ② for this reason
- ③ as a matter of fact
- ④ in addition to this

- ① however
- ② therefore
- ③ for instance
- ④ on the other hand

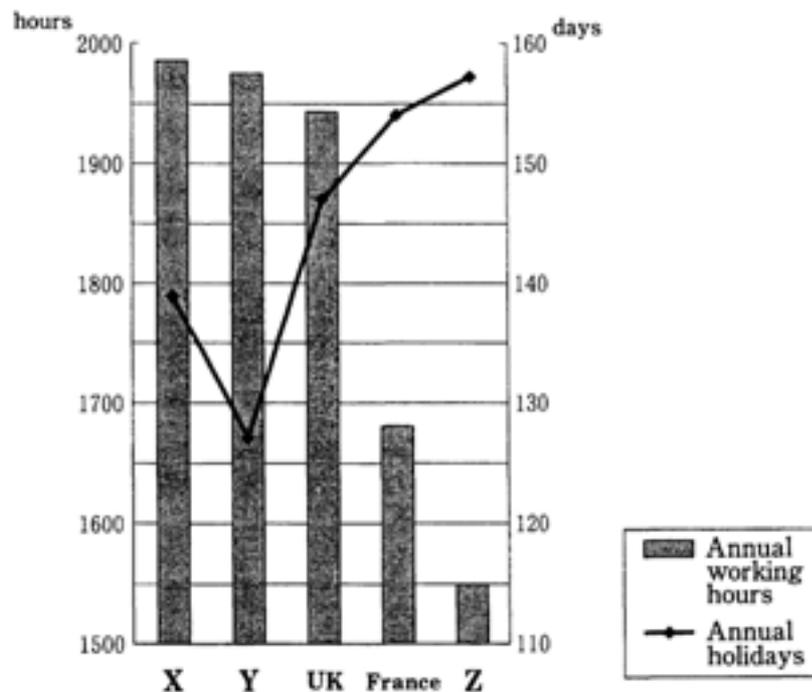
第4問 次の文章とグラフを読み、下の問い(A・B)に答えよ。(配点 35)

How hard do Japanese people work? Do they have enough free time to enjoy family life, travel, or hobbies? We have compared the working hours and holidays of workers in the manufacturing industry of five nations for the year 1995.

There are national holidays in each of the five countries, and people often enjoy three-day weekends. Japan leads this group in terms of the number of national holidays per year, with 15 days on its calendar such as Children's Day and Culture Day.

What if we look at the total number of holidays per year? This would include not only national holidays but also weekends and paid vacation. Germany led the five nations in the total number of annual holidays with 157 days. France came in second with 154 days. The United Kingdom (UK) was third, followed by the United States of America (USA). Japan came in last

Annual Working Hours and Holidays (1995)



(Data: Japanese Ministry of Labor, *White Paper on Labor*, 1997)

with the smallest number of holidays.

Some people say that the Japanese are taking more time off than they used to, and that the number of Japanese vacationing overseas is high. This may be true, but statistics show that, in 1995 at least, Western Europeans and Americans took more holidays than the Japanese did.

Most Western Europeans take about four weeks of paid vacation a year, while the Japanese and Americans usually take much less than that. A problem that seems to be unique to the Japanese is that only a small percentage of the paid vacation offered is actually taken. There are a lot of Japanese who do not feel comfortable about taking time off while their fellow workers have to continue working. Some Japanese business leaders, however, have encouraged workers to take more paid vacations in an effort to reduce international criticism that the Japanese work too hard.

Surprisingly, Japan came in second to the USA in terms of the average number of annual working hours at 1,975 hours in 1995. The United Kingdom came in third at 1,943 hours, followed by France at 1,680 hours. Germans worked the fewest hours that year. It is interesting that research such as this shows that facts do not always match the impression most people have.

A グラフのX, Y, Zに対応する国名の組合せが、その順に示されている。最も
適当な配列のものを、次の①～⑥のうちから一つ選べ。 33

- ① Germany — Japan — USA
- ② Japan — USA — Germany
- ③ Japan — Germany — USA
- ④ USA — Japan — Germany
- ⑤ USA — Germany — Japan
- ⑥ Germany — USA — Japan

B 次の問い(問1～4)の ～ に入れるのに最も適当なものを、それぞれ下の①～④のうちから一つずつ選べ。

問1 Japan leads the five nations in the number of .

- ① paid holidays
- ② annual holidays
- ③ national holidays
- ④ working hours

問2 Japan is different from the other nations in that .

- ① the Japanese have only about half as many days off as Western Europeans
- ② many Japanese workers hesitate to take paid vacations
- ③ most Japanese go abroad to enjoy their holidays far away from work
- ④ most of the holidays offered to workers each year are national holidays

問3 Among the five countries, are similar in that their workers receive more vacation time and spend fewer hours at work than the others.

- ① the UK and France
- ② Germany and Japan
- ③ Japan and the USA
- ④ France and Germany

第5問 次の会話について、下の問い(A～D)に答えよ。(配点 32)

Yuki and Gerry are making a homepage for their class on-line newspaper, "The Class e-Times."

Yuki: OK, let's check what we've got so far. We've decided to have one main story and one short story, right?

Gerry: Right. And what about pictures? Should we have one for each story?

Yuki: (1) Maybe it would be too much. How about just for the main story?

Gerry: That sounds good. Now, what will our stories be? We could do one about the students who visited from Korea. Maybe we can use one of the photos they sent us.

Yuki: Mmm, we could do that for the short story. I think I'd rather have the bus tour we took to Kyoto as the main story, though. (1)It was a lot of fun, and I'm sure people still remember the trip very clearly.

Gerry: (2) What about that great story you wrote about the trip for the homework assignment? We can use that as it is, if you could type (2)it into the computer. I can't write about it myself because I missed it.

Yuki: Oh, I remember. You were sick, weren't you?

Gerry: That's right. I wish I could have gone there.

Yuki: It was great. Kinkakuji was beautiful.

Gerry: You must have seen a lot of women in kimono.

Yuki: Not really. And even when we went to Gion later, we only saw a couple of them.

Gerry: (3) Have we got photos of them to go with the story?

Yuki: Yes, I've got a few here on my desk, but we decided we'd have just one picture. Do you think (3)it should go with the story about the Korean students or this one?

Gerry: I think it would be better to use one of your Kyoto photographs. Those pictures came out clearer.

Yuki: You're right. Look, I have this beautiful one of a woman in kimono, and these, here, of Kinkakuji.

Gerry: I like the first one. Can we use that?

Yuki: OK. The only thing now is to decide where to put each story.

Gerry: I've got an idea. We could have the main story in the left column, with the picture next to (4)it at the top, and the Korea story under the picture.

Yuki: I like that. It's got good balance. So, it looks like we're almost there. I'm beginning to see the light at the end of the tunnel.

Gerry: Yes. We're getting a pretty clear picture of what the homepage will look like, aren't we?

A 空欄(1)～(3)に入る三つの表現が、順不同で次のA～Cに示されている。意味の通る会話にするのに最も適当な配列のものを、下の①～⑥のうちから一つ選べ。 38

- A. Oh, really?
 B. That's a good idea.
 C. I'm not so sure about that.

- | | | |
|---------|---------|---------|
| ① A—B—C | ② A—C—B | ③ B—A—C |
| ④ B—C—A | ⑤ C—A—B | ⑥ C—B—A |

第6問 次の文章を読み、下の問い(A・B)に答えよ。(配点 45)

My confidence as a swimmer started to disappear the day Angela moved to our small town. At the time, some members of the town's swimming club, myself included, were preparing for the National Championships, which were just six months away. I had always been the best, and everyone thought that I would be chosen for the relay race. But now I had competition. There was only one place for the butterfly on the relay team, and we both wanted it.

For two weeks it was awful. Angela was always the star. She was faster than I, and her form was better, too. I was jealous and scared. My chances of being selected were disappearing fast. My fear caused me to be unfriendly to Angela. I refused to speak to her and never said anything good about her.

One day, however, our coach called me over and said, "Kate, I've got something to say to you. Your attitude is hurting your performance. I know you can change that. I'd like you to think about it."

When I arrived at the pool the next morning, I thought about what he had said as I was going through my warm-up. Angela and I were going to compete that morning, and only eight girls would enter the finals.

My thoughts were interrupted when a voice said, "Nervous?" It was Angela. "I don't like to talk before a race," I replied coldly.

"I get nervous, too," she said. Her voice didn't have the anger of mine, which surprised me quite a bit.

Angela and I competed in our separate trial races. I jumped into the pool and swam like a flying fish. My mind was clear, and I could think about only one thing: swimming well.

When the races were over, the judges announced that both of us were among the lucky eight who would be competing in the finals. Despite this good news, I noticed that Angela was sitting sadly alone. This puzzled me, but I thought that I knew how she felt and tried to be friendly to her.

"I don't talk before races, but I do talk *after* them. Sometimes it helps," I said.

Angela was silent for a while, but then she said, "I'm great in practices, but in competitions I just can't do well. It's like this all the time. I'm so worried about the finals."

Now I felt really bad. I realized how horrible I had been to Angela. I wanted to help her. I wanted to show her that I was sorry for my behavior of the past two weeks.

"Listen, I have an idea," I said. "Why don't we help each other prepare for the final race? We have two weeks to work on things."

"Good idea," said Angela.

For the next two weeks Angela and I worked together. I taught her how to deal with stress and how to train her muscles. She helped me with my form, and at the end of those two weeks we were the best of friends and respected each other as swimmers.

The day of the final race came and when the starter pistol was fired, I swam off like a dolphin. I thought about nothing but winning, but just before I reached the finish line, I thought of Angela and looked over into her lane.

As I was climbing out of the pool, I said to myself, "Oh no, what have I done?" I thought that I had dropped behind Angela and lost the race when I looked in her direction.

While we were waiting for the official announcement, the coach came running over to us. "Congratulations, girls! It was close, but Angela has won, and so have *you*, Kate!"

"What do you mean?" I asked.

"Well, Angela will be in the relay, but you swam so fast, Kate, that you, too, have won."

"Really?" I screamed. "I don't understand."

"I'll explain. The 21st Century Swimsuit Company is giving both of you their Future Swimmers Scholarship this year."

"Wow, I suppose that I'm a double winner: I got a scholarship and made a friend, too."

"You're not the only one," Angela remarked.

A 次の問い(問1～5)に対する答えとして最も適当なものを、それぞれ下の①～④のうちから一つずつ選べ。 43 ～ 47

問1 How did Kate feel after Angela joined the swimming club? 43

- ① Happy.
- ② Excited.
- ③ Sad.
- ④ Anxious.

問2 Why did the coach talk to Kate before the trial races? 44

- ① He thought she should be independent.
- ② He wanted to encourage her to swim better.
- ③ He wanted her to be afraid of Angela.
- ④ He thought she should support Angela.

問3 Why did Kate want to help Angela? 45

- ① She wanted Angela to win.
- ② The coach insisted that she do so.
- ③ Angela was a newcomer.
- ④ She understood how Angela felt.

問4 Why was Kate disappointed immediately after the final race? 46

- ① She thought she would not be going to the Nationals.
- ② She thought Angela had lost.
- ③ She had decided to give up swimming.
- ④ She found that Angela was not her friend.

Appendix II

The Society for Testing English Proficiency (STEP) performance descriptors for grades 1 – 5

1st Grade (Japanese college graduate level)

Able to conduct all aspects of daily life in English. Capable of conducting business activities in areas where only English is spoken. Able to understand a broad range of English and to express personal views with fluency.

The successful examinee is:

- (1) able to sufficiently understand TV/radio broadcasts, lectures, speeches, etc., and to orally express intentions (make a speech, debate, negotiate, etc.).
- (2) able to read high-level materials (newspaper editorials and commentaries, specific magazine articles, etc.) and understand them thoroughly.
- (3) able to clearly express ideas (including business matters) in writing.

Pre-1st Grade (Japanese college/junior college intermediate level)

Able to conduct daily life in English-speaking areas, with the exception of specialized fields. Although problems may remain in understanding natural English or expressing oneself perfectly, is able to grasp major points in order to communicate.

The successful examinee is:

- (1) able to converse about ordinary matters of daily life and relatively specialized matters of a personal nature, and able to grasp the gist of general speeches, lectures, broadcasts, etc.
- (2) able to read high-level materials (newspaper articles, general magazine articles, etc.) and understand the essential information.
- (3) able to write about personal ideas and opinions.

2nd Grade (Japanese high school graduate level)

Able to understand and use English well enough for everyday needs and situations. Able to communicate if visiting or living overseas.

The successful examinee is:

- (1) able to converse about basic matters of daily life (make simple explanations, conduct simple business by telephone, etc.).
- (2) able to read materials related to basic daily life (general newspaper and magazine articles, pamphlets, instructions, etc.).
- (3) able to write about basic matters of daily life.

Pre-2nd Grade (Japanese high school intermediate level)

Possesses the basic ability to cope with simple matters concerning daily life in English. Able to communicate in limited circumstances if visiting or living overseas. Able to exchange general information about oneself in simple English.

The successful examinee is:

- (1) able to converse about common daily topics (make simple conversation on the telephone, leave messages, etc.).
- (2) able to read materials about common everyday topics (simple news articles, letters, pamphlets, etc.).
- (3) able to write about common everyday topics.

3rd Grade (Japanese junior high school graduate level)

Able to understand elementary English on matters related to oneself, family, and friends. Able to handle easy daily conversation.

The successful examinee is:

- (1) able to hold a simple conversation about matters of daily life (make and reply to greetings, introduce someone, make purchases, etc.).
- (2) able to read simple materials (letters, signboards, notices, etc.).
- (3) able to write about oneself in elementary English.

4th Grade (Japanese junior high school intermediate level)

Capable of understanding basic English on matters concerning one's family and school life. Able to attempt communication to a limited extent.

The successful examinee is:

- (1) able to converse using set words, phrases, and sentences (give simple greetings, answer simple questions, etc.).
- (2) able to read simple passages (grasp the main points of short passages, letters, etc.).
- (3) able to write basic English.

5th Grade (Japanese junior high school beginning level)

Able to understand and express oneself in English with limited, fixed expressions.

The successful examinee is:

- (1) able to communicate through the use of basic words and phrases (use daily greetings, say yes and no correctly, etc.).
- (2) able to read at a beginning level (distinguish letters of the alphabet, understand basic points in a simple sentence, etc.).
- (3) able to write about simple matters using beginning-level English.

Retrieved from the STEP Test official website:

www.eiken.or.jp/english/evaluate/index.html

Appendix III

New speaking test and sample materials

Part One

Task: Learners introduce each other and interact with the interlocutor/assessor.

Timing: 4– 5 minutes

Interlocutor/assessor – use the following script:

Good afternoon, please sit down.

How are you Learner 1? And you, Learner 2?

Are you ready to start the test now? OK let's begin.

Learner 1, could I have your full name, please?

And how do you spell your surname?

(Ask for clarification. Candidate offers confirmation or repeats with repair if necessary).

Thank you.

And, Learner 2, could I have your full name, please?

How do you spell your surname?

(Ask for clarification. Candidate offers confirmation or repeats with repair if necessary).

Thank you.

Now, Learner 1, I'd like you to tell me something about Learner 2 – for example, where (s)he lives, what his/her hobbies are, which school club (s)he belongs to.

Thank you.

Learner 2, could you introduce Learner 1 now, please?

Thank you.

Interlocutor/assessor – for the next stage of the interaction, select one of the following cards and ask the questions to Learner 1. Then, select another card and ask the questions to Learner 2.

Option 1

What kind of place is _____ ?
How many are there in your family?
Describe your (brother / sister / mother / father) for me.
Thank you.

Option 2

Have you ever visited any other countries?
How long did you stay in _____ ?
What did you enjoy most about your visit?

Option 3

How do you get to school every day?
What's your favourite subject?
What do you like about it?
(Back-up: Do you enjoy studying English? / What do (don't) you like about it?

Option 4

How long have you been a student at this school?
When do you graduate?
What are your plans after graduating?
Thank you.

Part Two

Task: Learners are given copies of the same map, but each has different places labelled (thus creating an information-gap). The speaker (Learner 1) describes a route along the map. The listener (Learner 2) traces the route and marks the final destination, in pencil. Roles are then reversed.

Timing: 3 – 4 minutes

Interlocutor – use the following script:

Now, for this part of the test I'm going to give each of you a map and instruction card. Please read them carefully for a moment, but do not show the maps or the cards to each other.

Give out the materials and allow 30 seconds.

All right? Learner 2, you are playing the part of the tourist so I'd like you to start the conversation. When you are ready, please begin.

Allow the students enough time to complete the task. Prompt only if necessary.

Thank you.

Now, I'd like to give you another card but this time, Learner 2, you are going to give directions to Learner 1

Give out the second set of instruction cards and allow 30 seconds.

All right? Learner 1, this time you are playing the part of the tourist, so I'd like you to start this conversation. When you are ready, please begin.

Allow the students enough time to complete the task. Prompt only if necessary.

Thank you.

Sample instruction cards (places vary according to the card selected):

Card A

The situation:

A tourist (played by your partner) is lost, and approaches you outside the station to ask for directions.

What you have to do:

Use the map and give directions. Answer any other questions the tourist asks.

Card B

The situation:

You have just arrived at the station and approach a stranger (played by your partner) for directions to **the Grand Hotel**.

What you have to do:

Ask for directions and mark the route on your map. You may ask the stranger to repeat or clarify anything you are not sure about.

Part Three

Task: Two pictures are given to the learners, who must then interact with each other in order to discover the similarities and differences.

Timing: 3 minutes

Interlocutor – use the following script:

For the final part of the test, I'm going to give you both a picture. Again, I'd like you to look at them carefully, but do not show them to each other.

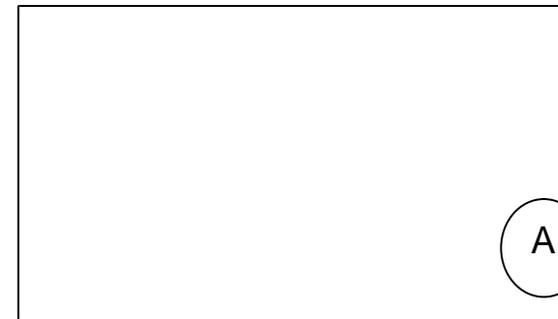
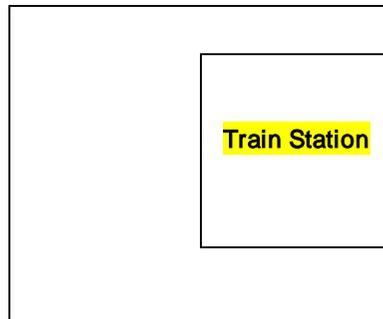
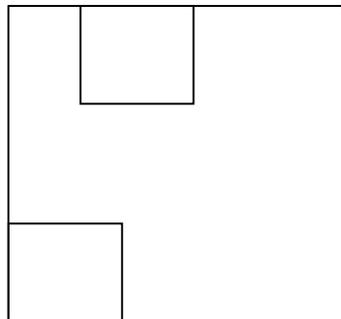
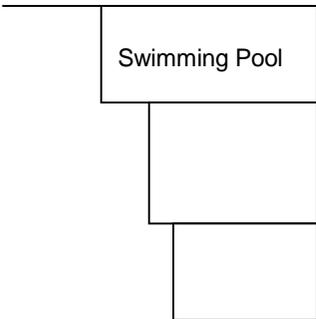
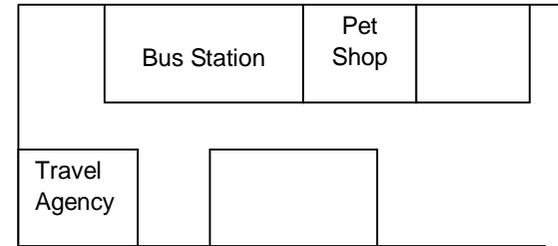
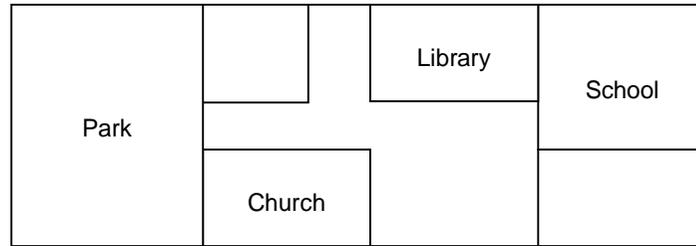
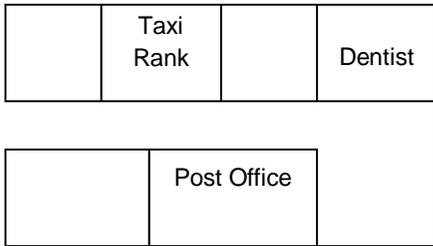
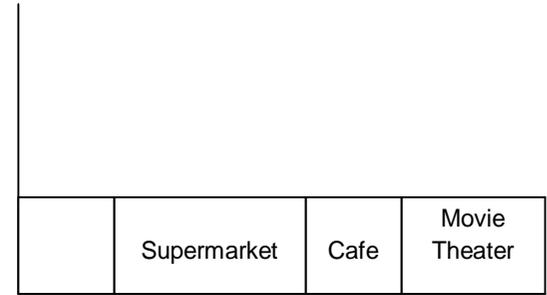
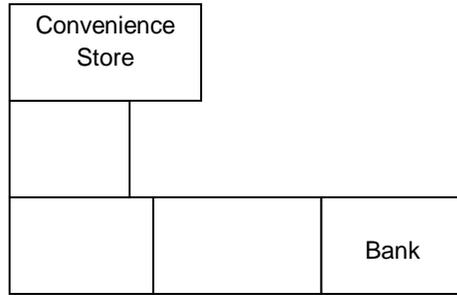
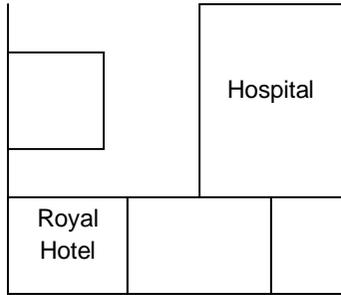
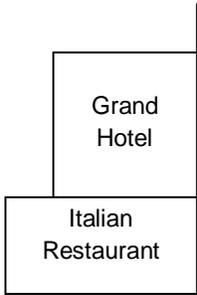
Give out the pictures and allow 30 seconds.

Now, the pictures you have been looking at are similar, but there are some differences. I'd like you both to talk about your pictures and find out how they are similar and how they are different.

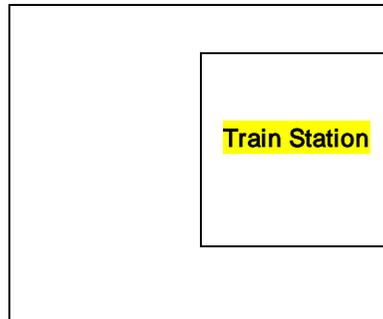
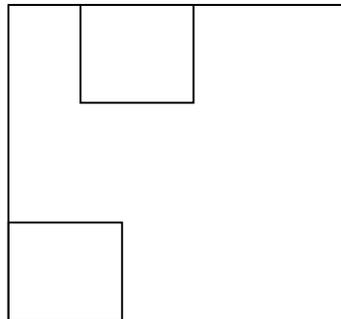
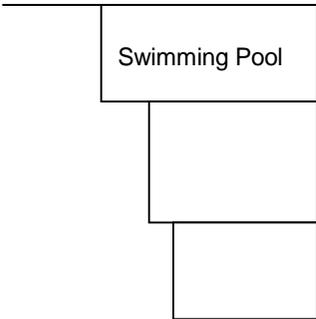
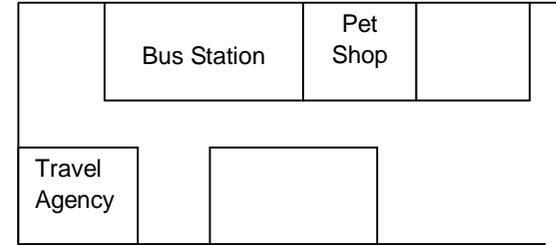
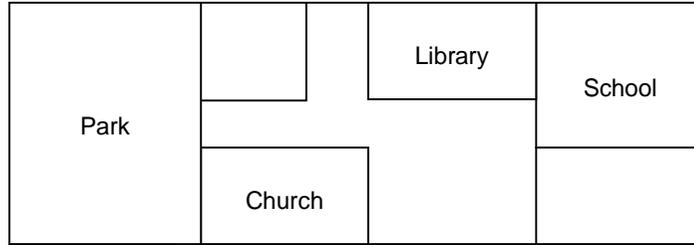
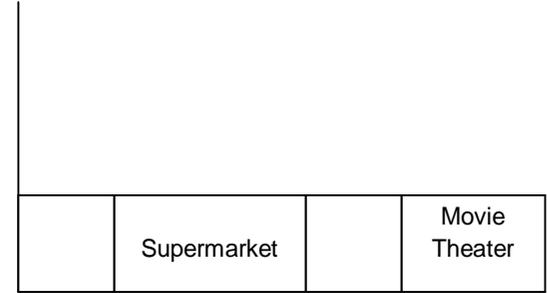
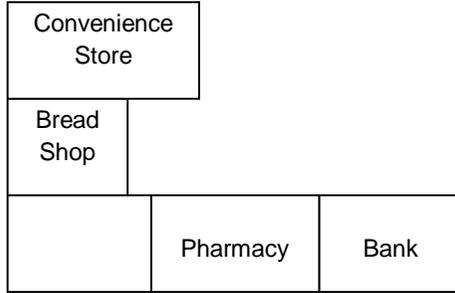
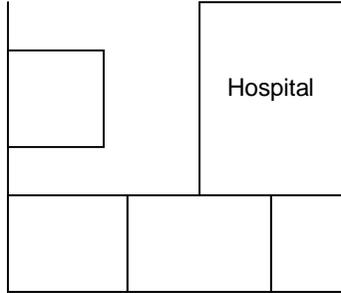
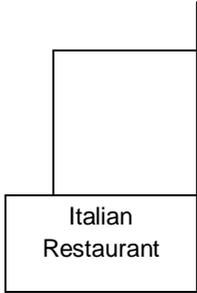
OK? Please begin

Allow the students enough time to complete the task. Prompt only if necessary.

Thank you. That is the end of the test.

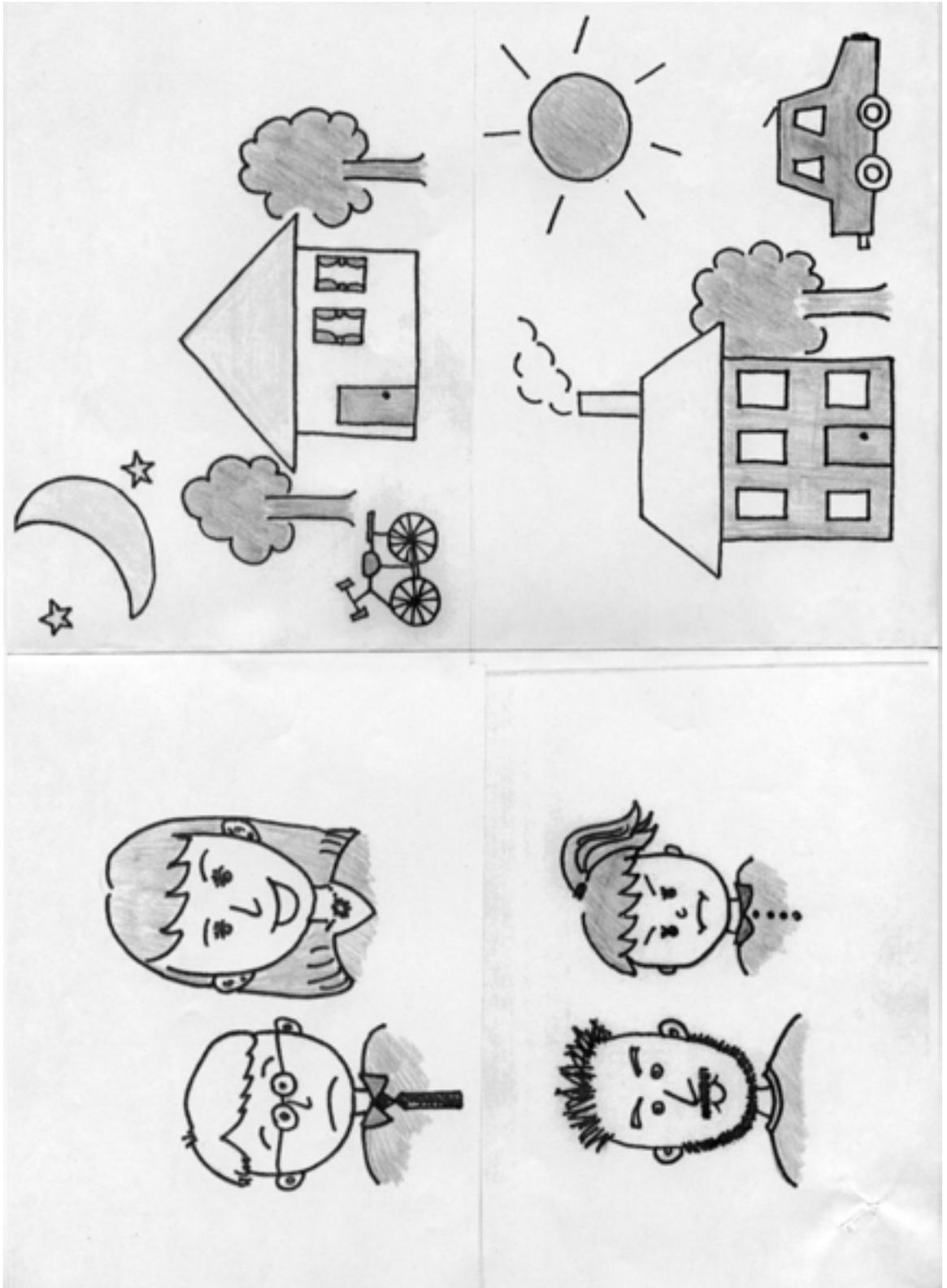


A



B

Sample materials for part three



Appendix IV
Analytic marking scheme for the speaking test

Score	Pronunciation	Vocabulary & Grammar	Fluency & Spontaneity	Communication
7 - 8	The candidate is easily understood and demonstrates accurate pronunciation. No evidence of katakana syllabary.	The candidate uses a wide range of appropriate vocabulary and structures.	The candidate responds fully and is able to develop the interaction with considerable ease.	The candidate fulfils all of the task requirements without the need for prompting from the assessor. Never reverts to L1 when “thinking aloud” or interacting.
5 - 6	The candidate is understood, although a few obvious pronunciation errors occur which occasionally affect successful communication. Little evidence of the katakana syllabary.	The candidate uses a good range of vocabulary and structures although a few errors do occur. Is able to adopt suitable coping strategies in order to tackle unknown vocabulary / structures.	The candidate responds well but hesitations and pauses sometimes affect the listener’s understanding of the message. Occasionally develops the interaction beyond the minimal requirements.	The candidate fulfils most of the task requirements without the need for prompting from the assessor. Sometimes reverts to L1 when “thinking aloud”.
3 - 4	The candidate is understood with some strain on the part of the listener. Obvious pronunciation errors occur which affect successful communication. Often relies upon the katakana syllabary.	The candidate uses an adequate range of vocabulary and structures although some are inappropriate for the task. Struggles to implement coping strategies to tackle unknown vocabulary / structures.	The candidate is able to respond but hesitations and pauses tend to strain the listener. Unable to develop the interaction beyond the minimal requirements.	The candidate fulfils some of the task requirements, but prompting from the assessor is sometimes required. Sometimes reverts to L1 when “thinking aloud” and interacting.
1 - 2	It is difficult to understand the candidate. Obvious pronunciation errors occur which frequently affect successful communication. Complete dependence upon the katakana syllabary.	The candidate uses a limited range of vocabulary and structures which are often inappropriate for the task.	The candidate is often unable to respond and frequently fails to meet the minimal requirements of the task.	The candidate is unable to fulfil the task requirements without frequent prompting from the assessor. Often dependent on L1 to achieve successful communication.
0	Insufficient amount of language produced for assessment			

Appendix VI

Results from the classroom observation sessions using the COLT (Part A)

Observation Scheme

COLT CATEGORIES		TEACHER A		TEACHER B	
		SESSION 1	SESSION 2	SESSION 3	SESSION 4
PARTICIPANT ORGANISATION	T<->S/C	63.41	94.00	43.59	57.50
	S<->S/C	0.00	0.00	5.13	0.00
	(Choral & T<->S/C)	0.00	0.00	12.82	0.00
	Group	0.00	0.00	0.00	0.00
	Individual	36.59	6.00	38.46	15.00
	(Individual & S<->S/C)	0.00	0.00	0.00	27.50
CONTENT	Procedure	0.00	0.00	0.00	0.00
	Form	100.00	100.00	100.00	100.00
	Function	0.00	0.00	0.00	0.00
	Discourse	0.00	0.00	0.00	0.00
	Sociolinguistics	0.00	0.00	0.00	0.00
	Others	0.00	0.00	0.00	0.00
CONTENT CONTROL	Teacher/Text	100.00	100.00	100.00	100.00
	Teacher/Text/Student	0.00	0.00	0.00	0.00
	Student	0.00	0.00	0.00	0.00
STUDENT MODALITY	Listening	0.00	20.00	5.13	5.00
	(Listening &	9.76	46.00	25.64	17.50
	(Listening &	53.66	22.00	12.82	0.00
	(List./Speaking/Reading)	0.00	0.00	12.82	0.00
	Speaking	0.00	0.00	0.00	0.00
	Reading	0.00	0.00	0.00	7.50
	(Reading & Writing)	0.00	0.00	0.00	15.00
	Writing	36.59	12.00	43.59	55.00
MATERIALS	Minimal Text	100.00	100.00	100.00	100.00
	Extended Text	0.00	0.00	0.00	0.00
	Audio	0.00	0.00	0.00	0.00
	Visual	0.00	0.00	0.00	0.00
	L2 – NNS	100.00	100.00	100.00	100.00
	L2 – NS	0.00	0.00	0.00	0.00
	L2 – NSA	0.00	0.00	0.00	0.00
	Student-made	0.00	0.00	0.00	0.00

Appendix VII

Teacher questionnaire

Please answer all questions as completely as possible.

If you wish to add any further comments, please use the space provided after each question.

(1) How long have you been an English teacher?

years

(2) Which courses and grades are you currently teaching?

(Please tick the appropriate boxes).

<input type="checkbox"/>	<u>English I / II</u>
<input type="checkbox"/>	1 st grade
<input type="checkbox"/>	2 nd grade
<input type="checkbox"/>	3 rd grade

<input type="checkbox"/>	<u>English IG / IIG</u>
<input type="checkbox"/>	1 st grade
<input type="checkbox"/>	2 nd grade
<input type="checkbox"/>	3 rd grade

<input type="checkbox"/>	<u>English IR / IIR</u>
<input type="checkbox"/>	1 st grade
<input type="checkbox"/>	2 nd grade
<input type="checkbox"/>	3 rd grade

<input type="checkbox"/>	<u>Oral Communication I / II</u>
<input type="checkbox"/>	1 st grade
<input type="checkbox"/>	2 nd grade
<input type="checkbox"/>	3 rd grade

<input type="checkbox"/>	<u>English Special Course</u>
<input type="checkbox"/>	1 st grade
<input type="checkbox"/>	2 nd grade
<input type="checkbox"/>	3 rd grade

Comments:

(3) Which language do you use in the classroom? (Please tick one box).

<input type="checkbox"/>	English only
<input type="checkbox"/>	English with occasional Japanese explanation
<input type="checkbox"/>	Half English and half Japanese
<input type="checkbox"/>	Mainly Japanese

Comments:

(4) Look at the following patterns of classroom organization. How often do you use each one?

(Please rate your response on a 4 point scale where 1=never; 2=sometimes; 3=often; 4=always. Circle your answers.)

- | | | | | |
|------------------------------------|---|---|---|---|
| (a) teacher-fronted lecture | 1 | 2 | 3 | 4 |
| (b) pair work | 1 | 2 | 3 | 4 |
| (c) group work | 1 | 2 | 3 | 4 |
| (d) individual (student) desk work | 1 | 2 | 3 | 4 |
| (e) student presentations | 1 | 2 | 3 | 4 |

Comments:

(5) How important do you think it is for students to attend a juku or yobiko in order to pass university entrance examinations?

(Please rate your response on a 4 point scale where 1=not at all important; 2=not so important; 3=important; 4=very important. Circle your answers.)

1 2 3 4

Comments:

(6) How important do you think it is to test students' writing and speaking ability?

(Please rate your response on a 4 point scale where 1=not at all important; 2=not so important; 3=important; 4=very important. Circle your answers.)

1 2 3 4

Comments:

(7) In Japan, how important do you think it is for students to develop communicative ability in English?

(Please rate your response on a 4 point scale where 1=not at all important; 2=not so important; 3=important; 4=very important. Circle your answers.)

1 2 3 4

Comments:

(8) Who do you think should teach English listening and speaking skills?

(Please rate your responses on a 4 point scale where 1=strongly disagree; 2=disagree; 3=agree; 4=strongly agree. Circle your answers.)

- | | | | | |
|--|---|---|---|---|
| (a) Native speaker teachers of English | 1 | 2 | 3 | 4 |
| (b) Japanese teachers of English (JTE's) | 1 | 2 | 3 | 4 |
| (c) One JTE and one native speaker (team teaching) | 1 | 2 | 3 | 4 |

Comments:

(9) Why do you think it is important for your students to study English?

(Please rate your responses on a 4 point scale where 1=strongly disagree; 2=disagree; 3=agree; 4=strongly agree. Circle your answers.)

- | | | | | |
|--|---|---|---|---|
| (a) to graduate from high school | 1 | 2 | 3 | 4 |
| (b) to pass university entrance examinations | 1 | 2 | 3 | 4 |
| (c) to achieve good scores on the Seibu-moshi
and Kenka-Issei tests | 1 | 2 | 3 | 4 |
| (d) to pass the Eiken test | 1 | 2 | 3 | 4 |
| (e) to travel abroad | 1 | 2 | 3 | 4 |
| (f) to study/work abroad | 1 | 2 | 3 | 4 |
| (g) for future employment | 1 | 2 | 3 | 4 |
| (h) to study English at university | 1 | 2 | 3 | 4 |
| (i) to communicate with English-speaking
visitors who come to Japan | 1 | 2 | 3 | 4 |
| (j) because it is educationally and culturally
desirable to have knowledge of English | 1 | 2 | 3 | 4 |
| (k) to satisfy their parents' expectations | 1 | 2 | 3 | 4 |

Comments:

(10) Which of the following areas do you think are most important for Japanese learners of English?

(Please rate your responses on a 4 point scale where 1=strongly disagree; 2=disagree; 3=agree; 4=strongly agree. Circle your answers.)

- | | | | | |
|-----------------------------|---|---|---|---|
| (a) Listening comprehension | 1 | 2 | 3 | 4 |
| (b) Writing | 1 | 2 | 3 | 4 |
| (c) Reading comprehension | 1 | 2 | 3 | 4 |
| (d) Speaking | 1 | 2 | 3 | 4 |
| (e) Grammar | 1 | 2 | 3 | 4 |
| (f) Vocabulary | 1 | 2 | 3 | 4 |
| (g) Spelling | 1 | 2 | 3 | 4 |
| (h) Pronunciation / stress | 1 | 2 | 3 | 4 |

Comments:

(11) Which of the following language resources do you think should be used for teaching English to Japanese high school students?

(Please rate your responses on a 4 point scale where 1=strongly disagree; 2=disagree; 3=agree; 4=strongly agree. Circle your answers.)

- | | | | | |
|--|---|---|---|---|
| (a) listening exercises using authentic monologues / dialogues | 1 | 2 | 3 | 4 |
| (b) listening exercises using adapted monologues / dialogues
(i.e. materials made specially for EFL learners) | 1 | 2 | 3 | 4 |
| (c) recommended textbooks | 1 | 2 | 3 | 4 |
| (d) authentic video (e.g. English movies/TV shows) | 1 | 2 | 3 | 4 |
| (e) adapted video made specially for EFL learners | 1 | 2 | 3 | 4 |
| (f) English music recordings | 1 | 2 | 3 | 4 |
| (g) past exam papers | 1 | 2 | 3 | 4 |
| (h) teacher-prepared worksheets | 1 | 2 | 3 | 4 |
| (i) language learning websites/e-mail | 1 | 2 | 3 | 4 |
| (j) fiction/non-fiction EFL graded reader books | 1 | 2 | 3 | 4 |
| (k) English newspapers / magazines | 1 | 2 | 3 | 4 |
| (l) English word games or puzzles | 1 | 2 | 3 | 4 |

Comments:

(12) Which of the following teaching strategies/activities do you think should be used for teaching English to Japanese high school students?

(Please rate your responses on a 4 point scale where 1=strongly disagree; 2=disagree; 3=agree; 4=strongly agree. Circle your answers.)

- | | | | | |
|---|---|---|---|---|
| (a) reading aloud (teacher) | 1 | 2 | 3 | 4 |
| (b) reading aloud (learner) | 1 | 2 | 3 | 4 |
| (c) silent reading | 1 | 2 | 3 | 4 |
| (d) role play/simulation | 1 | 2 | 3 | 4 |
| (e) information gap activities | 1 | 2 | 3 | 4 |
| (f) repetition/substitution drills | 1 | 2 | 3 | 4 |
| (g) copying words/sentences | 1 | 2 | 3 | 4 |
| (h) dictation | 1 | 2 | 3 | 4 |
| (i) letter writing | 1 | 2 | 3 | 4 |
| (j) creative writing (stories etc.) | 1 | 2 | 3 | 4 |
| (k) class discussion/debate | 1 | 2 | 3 | 4 |
| (l) multiple-choice grammar/vocabulary exercises | 1 | 2 | 3 | 4 |
| (m) multiple-choice reading comprehension exercises | 1 | 2 | 3 | 4 |
| (n) gap-filling/completion exercises | 1 | 2 | 3 | 4 |
| (o) translating English sentences/texts into Japanese | 1 | 2 | 3 | 4 |
| (p) translating Japanese sentences/texts into English | 1 | 2 | 3 | 4 |
| (q) speaking activities in pairs or groups | 1 | 2 | 3 | 4 |

Comments:

(13) Which of these English tests do you think offer an effective measure of a student's speaking and writing ability?

(Please rate your responses on a 4 point scale where 1=strongly disagree; 2=disagree; 3=agree; 4=strongly agree. Circle your answers – one in the 'Speaking' column; one in the 'Writing' column.)

<u>SPEAKING</u>					<u>WRITING</u>			
1	2	3	4	(a) Seibu moshi/Kenka Issei	1	2	3	4
1	2	3	4	(b) The Center test	1	2	3	4
1	2	3	4	(c) The Eiken test	1	2	3	4
1	2	3	4	(d) National university entrance tests	1	2	3	4
1	2	3	4	(e) Local Public university tests	1	2	3	4
1	2	3	4	(f) Private university tests	1	2	3	4
1	2	3	4	(g) The TOEIC test	1	2	3	4

Comments:

(14) To what extent do the following factors influence your teaching?

(Circle your answers.)

	<u>INFLUENCE</u>			
	a little		a lot	
(a) professional training	1	2	3	4
(b) teaching experience	1	2	3	4
(c) MEXT syllabus	1	2	3	4
(d) past experience as a language learner	1	2	3	4
(e) textbooks	1	2	3	4
(f) the need to obtain satisfaction in teaching	1	2	3	4
(g) external examinations	1	2	3	4
(h) learners' expectations	1	2	3	4
(i) school's expectations	1	2	3	4
(j) parents' expectations	1	2	3	4

Comments:

(15) The 2006 Center Test will contain a listening section for the first time. Will this affect the way you prepare students for the test and if so, in what way?

Comments:

(16) What do you think of the current English tests taken by your students? (e.g. Seibu-moshi, Kenka-Issei, STEP, Center Test, University entrance examinations, TOEIC). What do you think are the positive and negative aspects of these tests? How, if at all, would you improve them?

Comments:

- END OF QUESTIONNAIRE-

Thank you very much for your help

Appendix VIII

Student questionnaire #1

Please answer all questions.

If you wish to add any further comments, please use the space provided after each question.

(1) Arrange these subjects in order of importance for you. (1=most important; 10=least important)

- | | |
|--------------------------|----------------------------------|
| <input type="checkbox"/> | Mathematics |
| <input type="checkbox"/> | English IIR |
| <input type="checkbox"/> | English IIG |
| <input type="checkbox"/> | Oral Communication II |
| <input type="checkbox"/> | Modern Japanese |
| <input type="checkbox"/> | Old Japanese |
| <input type="checkbox"/> | Japanese History / World History |
| <input type="checkbox"/> | Religion |
| <input type="checkbox"/> | Science |
| <input type="checkbox"/> | Health / Physical Education |

Comments:

(2) Where did you first start to learn English and how old were you?

(3) Do you have any English classes outside of school? (Yes / No)

If you answered 'yes', please give details (e.g. Is it a juku, yobiko or eikaiwa? How often do you attend? How long have you been studying there?)

Please read questions 4 – 6 and grade each part (a, b, c etc.) on a 4-point scale, where:

1 = strongly disagree

2 = disagree

3 = agree

4 = strongly agree

Please circle your answers.

(4) Why are you studying English?

- | | | | | |
|--|---|---|---|---|
| (a) I have to in order to graduate | 1 | 2 | 3 | 4 |
| (b) I want to do well on the university entrance exams | 1 | 2 | 3 | 4 |
| (c) I want to do well on the Seibu-moshi and Kenka-Issei | 1 | 2 | 3 | 4 |
| (d) I want to improve my Eiken level | 1 | 2 | 3 | 4 |
| (e) I want to visit English-speaking countries in the future | 1 | 2 | 3 | 4 |
| (f) I want to study / work in abroad in the future | 1 | 2 | 3 | 4 |
| (g) I want a job in which I can use English | 1 | 2 | 3 | 4 |
| (h) I want to study English at university | 1 | 2 | 3 | 4 |
| (i) I want to communicate with English-speaking
visitors to Japan | 1 | 2 | 3 | 4 |
| (j) I think it is interesting to learn a foreign language | 1 | 2 | 3 | 4 |

Comments:

(5) In English lessons, what do you think are the most important elements?

(a) Listening comprehension	1	2	3	4
(b) Writing	1	2	3	4
(c) Reading Comprehension	1	2	3	4
(d) Speaking	1	2	3	4
(e) Grammar	1	2	3	4
(f) Vocabulary	1	2	3	4
(g) Spelling	1	2	3	4
(h) Pronunciation / Stress	1	2	3	4

Comments:

(6) How helpful do you find the following activities in preparing for the examinations?

(a) listening exercises using taped monologues/dialogues	1	2	3	4
(b) studying with the textbook	1	2	3	4
(c) watching English movies/TV shows	1	2	3	4
(d) watching videos made specially for EFL learners	1	2	3	4
(e) listening to English music recordings	1	2	3	4
(f) studying past exam questions	1	2	3	4
(g) doing teacher-prepared worksheets	1	2	3	4
(h) using language learning websites	1	2	3	4
(i) reading fiction or non-fiction EFL books	1	2	3	4
(j) translating English sentences/texts into Japanese	1	2	3	4
(k) translating Japanese sentences/texts into English	1	2	3	4
(l) doing multiple-choice reading comprehension questions	1	2	3	4
(m) doing multiple-choice/blank-filling grammar/vocabulary exercises	1	2	3	4
(n) doing speaking activities in pairs or groups	1	2	3	4
(o) using English word games or puzzles	1	2	3	4

Comments:

(7) What do you think is the best way to prepare for an English test such as the Seibu Moshi or Center Test?

- | | | | | |
|------------------------------------|---|---|---|---|
| (a) in class with a teacher | 1 | 2 | 3 | 4 |
| (b) out of class by yourself | 1 | 2 | 3 | 4 |
| (c) out of class with friends | 1 | 2 | 3 | 4 |
| (d) with a private tutor at a juku | 1 | 2 | 3 | 4 |

Comments:

(8) Which of the learning activities listed in question 6 have you used when studying in your own time? (Write down the letter of the activity and add further comment if necessary)

– That is the end of the questionnaire –

Thank you very much for your help

Appendix IX

Student questionnaire #2

The following statements refer to the end-of-term speaking test you have just done in Oral Communication. Grade each one on a 4-point scale, where:

1 = strongly disagree

2 = disagree

3 = agree

4 = strongly agree

Write your answers in the brackets.

- (1) [] I enjoyed doing the speaking test with a partner.
- (2) [] Studying for the speaking test improved my English.
- (3) [] I made a greater effort to speak English in the weeks leading up to the test.
- (4) [] It was important to do well on the test.
- (5) [] It was possible to do well on the test without much preparation.
- (6) [] If the test was included as part of the Seibu Moshi or Center Test I would do more preparation.
- (7) [] I enjoyed practising for the test during class.
- (8) [] It was important to practise for the test during class.
- (9) [] There wasn't enough time to practise for the test.
- (10) [] It is more important to practice other language skills during class.
- (11) [] When not in class, it is better to study for a speaking test on your own.
- (12) [] When not in class, it is better to study for a speaking test with friends.

(13) [] I would have liked to have had access to more materials for practice out of class.

(14) [] It is difficult to study for a speaking test in your own time without a teacher.

(15) [] It wasn't necessary to study for the speaking test in my own time.

(16) [] In order to do well on the test it is necessary to memorize key phrases/vocabulary.

(17) [] It is only necessary to speak English during Oral Communication classes.

(18) [] I want to improve my English speaking skills.

(19) [] It isn't important for me to speak English, so speaking shouldn't be tested.

Comments:

– That is the end of the questionnaire –

Thank you very much for your help

Bibliography

- Alderson J C & D Wall (1993) 'Does washback exist?' **Applied Linguistics** 14/2: 115 – 129
- Andrews S (2004) 'Washback and Curriculum Innovation' in Cheng L, Y Watanabe & A Curtis (Eds.)
- 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 (1999) **Washback in Language Testing. TOEFL Monograph Series, MS 15**: Princeton, NJ: Educational Testing Service
- Brown G & G Yule (1983) **Teaching the Spoken Language** Cambridge: CUP
- Brown J D (1995) 'English Language Entrance Examinations in Japan: Problems and Solutions' **On JALT '95: Proceedings of the JALT 1995 Conference** Tokyo: Japan Association for Language Teaching
- Brown J D (2001) **Using Surveys in Language Programs** Cambridge: CUP
- Brown J D (2002) 'Extraneous variables and the washback effect' **Shiken: JALT Testing & Evaluation SIG Newsletter** 6/2: 12 – 15 [available on-line]
www.jalt.org/test/bro_14.htm [last accessed 04.09.05]
- Brown J D & S O Yamashita (Eds.) (1995) **Language Testing in Japan** Tokyo: Japan Association for Language Teaching

- Brown J D & S O Yamashita (1995) 'English language entrance examinations at Japanese universities: 1993 and 1994' in Brown J D & S O Yamashita (Eds.)
- Brumfit C J & K Johnson (1979) **The Communicative Approach to Language Teaching**
Oxford: OUP
- Burrows C (2004) 'Washback in Classroom-Based Assessment: A Study of the Washback Effect in the Australian Adult Migrant English Program' in Cheng L, Y Watanabe & A Curtis (Eds.)
- Canale M & M Swain (1980) 'Theoretical bases of communicative approaches to second language teaching and testing' **Applied Linguistics** Vol.1, No.1: 1 – 47
- Chapman D W & C W Snyder Jr (2000) 'Can high stakes national testing improve instruction: reexamining conventional wisdom' **International Journal of Educational Development** 20: 457 – 474
- Cheng L (1997) How does washback influence teaching? Implications for Hong Kong' **Language and Education** 11/1: 38 – 54
- Cheng L (1999) 'Changing assessment: Washback on teacher perspectives and actions' **Teaching and Teacher Education** 15: 253 – 271
- Cheng L & A Curtis (2004) 'Washback or backwash: A review of the impact of testing on teaching and learning' in Cheng L, Y Watanabe & A Curtis (Eds.)
- Cheng L, Y Watanabe with A Curtis (Eds.) (2004) **Washback in Language Testing: Research Contexts and Methods** New Jersey: LEA

Chomsky N (1965) **Aspects of the theory of syntax** Cambridge, Mass.: MIT Press

Clapham C & D Corson (Eds.) (1997) **Encyclopedia of Language and Education: Vol.7.**

Language Testing and Assessment Dordrecht: Kluwer Academic

Collins S (1999) 'Communicative Methodology: A Health Warning?' **Explorations in Teacher Education** 7/1 [on-line]

www.jalt.org/teach/Newsletter_files/Jun_1999/June_1999_contents.html

[last accessed 04.09.05]

Davies A (1968) **Language Testing Symposium: A Psycholinguistic Approach** London: OUP

Davies A (1978) 'Language testing' **Language Teaching and Linguistics Abstracts II**, 3/4: 145 – 159

Fulcher G (2000) 'The 'communicative' legacy in language testing' **System** 28/4: 483 – 497

Gates S (1995) 'Exploiting washback from standardized tests' in Brown J D & S O Yamashita (Eds.)

Gorsuch G (1998) 'Yakudoku EFL instruction in two Japanese high school classrooms: an exploratory study' **JALT Journal** 20/1: 6 – 32

Guest M (2005) 'Entrance exams – not what you think' article in **The Daily Yomiuri** (28/01/05)

Hughes A (Ed.) (1988) **Testing English for University Study** (ELT Document 127)

Modern English Publications

Hughes A (1988) 'Introducing a needs-based test of English language proficiency into an English-medium university in Turkey' in Hughes A (Ed.)

Hughes A (1989) **Testing for Language Teachers** Cambridge: CUP

Hymes D H (1972) 'On communicative competence' in Pride J B & J Holmes (Eds.)

Ingram E (1968) 'Attainment and Diagnostic Testing' in Davies A (Ed.)

Johnson R K (Ed.) (1989) **The Second Language Curriculum** Cambridge: CUP

Johnson R K (1989) 'A decision-making framework for the coherent language curriculum' in Johnson R K (Ed.)

Kellaghan T, G. F Madaus & P W Airasian (1982) **The Effects of Standardized Testing**

London: Kluwen, Nijhoff Publishing

Kitao K & S K Kitao (1996) 'Testing Communicative Competence' **The Internet TESL**

Journal Vol.II, No.5 [on-line] <http://iteslj.org/Articles/Kitao-Testing.html> [last accessed 04.09.05]

Lado R (1961) **Language Testing: The Construction and Use of Foreign Language**

Tests London: Longman

MacGregor L (1997) 'The Eiken Test: An Investigation' **JALT Journal** 19/1

Madaus G. F (1988) 'The influence of testing on the curriculum' in Tanner L N (Ed.)

McNamara T (2000) **Language Testing** Oxford: OUP

Messick S (1996) 'Validity and washback in language testing' **Language Testing** 13/3:
241 – 256

MEXT (Ministry of Education, Culture, Sports, Science and Technology) (1993) **The
Course of Study for Foreign Languages** [available on-line]
www.mext.go.jp/english/shotou/030301.htm [last accessed 04.09.05]

MEXT (Ministry of Education, Culture, Sports, Science and Technology) (2003) **Action
Plan to Cultivate “Japanese with English Abilities”** [available on-line]
http://www.mext.go.jp/b_menu/houdou/15/03/03033101/001.pdf [last accessed
04.09.05]

Morrow K (1979) 'Communicative language testing: revolution or evolution?' in
Brumfit C J & K Johnson (Eds.)

Norris-Holt J (2002) 'An Investigation of Japanese High School Students' Attitudes
Towards the Study of English' **Second Language Learning & Teaching** Vol.2
[on-line] www.usq.edu.au/opacs/sllt/2/Norris-Holt02.html [last accessed
04.09.05]

Pride J B & J Holmes (Eds.) (1972) **Sociolinguistics** Harmondsworth, UK: Penguin

Rea-Dickins P (1997) 'So, why do we need relationships with stakeholders in language
testing? A view from the UK' **Language Testing** 14/3: 304-314

- Reesor M (2003) 'Japanese attitudes to English: Towards an explanation of poor performance' **NUCB Journal of Language, Culture and Communication** Vol.5, No.2: 57 – 65 [available on-line]
<http://www.nucba.ac.jp/cic/pdf/njlcc052/06REESOR.PDF> [last accessed 04.09.05]
- Richards J C & T S Rodgers (1986) **Approaches and Methods in Language Teaching**
Cambridge: CUP
- Sakui K (2004) 'Wearing two pairs of shoes: language teaching in Japan' **ELTJ** 58/2: 155 – 163
- Saville N (2000) 'Investigating the impact of international language examinations'
University of Cambridge Local Examinations Syndicate (UCLES) Research Notes 2 [available on-line] http://www.cambridge-efl.org/rs_notes [last accessed 04.09.05]
- Snyder C W Jr, B Prince, G Johanson, C Odaet, L Jaji & M Beatty (1997) 'Exam Fervor and Fever: Case Studies of the Influence of Primary Leaving Examinations on Uganda Classrooms, Teachers and Pupils (Volume One)' **Advancing Basic Education and Literacy (ABEL) Project** Academy for Educational Development, Washington DC
- Shohamy E, S Donitsa-Schmidt & I Ferman (1996) 'Test impact revisited: Washback effect over time' **Language Testing** 13/3: 298 – 317

Spada N & M Frohlich (1995) **COLT Observation Scheme: Communicative Orientation of Language Teaching Observation Scheme Coding Conventions and Applications** Sydney: NCELTR

Tanner L N (Ed.) (1988) **Critical Issues in Curriculum: Eighty-seventh Yearbook of the National Society for the Study of Education** Chicago: University of Chicago Press

Taylor L (2000) 'Stakeholders in language testing' **University of Cambridge Local Examinations Syndicate (UCLES) Research Notes 2** [available on-line]
http://www.cambridge-efl.org/rs_notes [last accessed 04.09.05]

Underhill N (1987) **Testing Spoken Language: A handbook of oral testing techniques** Cambridge: CUP

Wall D (1997) 'Impact and washback in language testing' in Clapham C & D Corson (Eds.)

Wall D (2000) 'The impact of high-stakes testing on teaching and learning: can this be predicted or controlled?' **System** 28/4: 499 – 509

Wall D & J C Alderson (1993) 'Examining washback: The Sri Lankan impact study' **Language Testing** 10/1: 41 – 69

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

Wesdorp H (1982) **Backwash effects of language testing in primary and secondary education** Stichting Centrum voor onderwijsonderzoek van de Universiteit van Amsterdam