



Investigating the washback potential of an English speaking test in the Japanese university entrance exam context

Date: 9 July 2025

Author name and ORCID number: David Allen
(0000-0002-3275-331X)

Grant reference: AR-G-2023-03

www.britishcouncil.org/english-assessment/english-language-research

This report is brought to you by English Language Research, British Council

To cite this report, please use: Allen, D., (2025). Investigating the washback potential of the BCT-S. British Council. DOI: 10.57884/AZNR-RT90

Contents

Executive summary.....	4
Introduction.....	5
Literature review.....	7
1.1 Washback.....	7
1.2 Washback research.....	7
1.3 Hypothetical washback research design.....	9
The present study.....	11
2.1 The BCT-S.....	11
2.2 Research design.....	12
Methodology.....	14
3.1 Participants.....	14
3.2 Procedure.....	17
3.3 Data analysis.....	18
Results and discussion.....	20
4.1 Washback effects.....	20
4.2 Washback direction, intensity and mediating factors.....	22
4.3 Educational goals.....	23
4.4 Overlap with the educational system at the school.....	24
4.5 Proximity.....	26
4.6 Learner factors.....	26
4.7 Teacher factors.....	27
4.8 Suggestions for test design.....	28
4.9 Facilitating washback through a ToA.....	28
Appendices.....	39
Appendix 1: BCT-S test design.....	39
Appendix 2: BCT-S-related research presentations.....	40
Appendix 3: Survey content.....	42
Appendix 4: Interview topics, supporting research, and example questions.....	45
Appendix 5: Narrative summaries.....	53



Executive summary

Test developers have been urged to work for washback, that is, to design language assessments that bring about positive washback as a consequence of use. Moreover, recent impact-by-design approaches to test development and validation argue that washback research can be conducted at various stages of the test development process, not simply following implementation. In line with these ideas, the present study conducted a hypothetical washback study into the expected effects of an English speaking test designed for use in university admissions in Japan. Nine senior high-school teachers (Grades 10–12) were familiarised with the test and interviewed about the expected washback effects and mediating factors of those effects, while imagining that the test would be introduced nationwide as a high-stakes test. Findings revealed a range of expected washback effects that were mediated by factors including the educational goals of the institution, the overlap between the test and teaching at the school, proximity, and various student and teacher factors, such as English proficiency, motivation and willingness to speak. The findings were used to develop a theory of action that could be used to promote positive consequences and manage potential negative consequences, if the test were to be implemented as an admissions test.

Introduction

Consequences of test use are often stated to be the guiding concern of test developers (e.g. Bachman & Palmer, 2010; Chalhoub-Deville & O'Sullivan, 2020; Kane, 2013). One of the most important consequences is washback, which is typically defined as the effect that a test has on teaching and learning (for example Hughes, 2003). Washback, and its superordinate term 'test impact', are typically seen as relating to a test's consequential validity (for example Weir, 2005), and washback research has traditionally been conducted once a test has been implemented. However, while this 'after-the-fact' approach to research has undoubtedly been the norm, researchers have long been urged to *work for washback* (Swain, 1985), and present-day systems for managing test impact (e.g. Saville & Khalifa, 2016; Saville & Séguis, 2025) view washback research as integral to all stages of the test development cycle, from conception to implementation. Concurrent with the ethos of working for washback, the present study investigates the potential washback expected from the introduction of a test of English speaking proficiency into the university entrance exam system in Japan.

The use of English exams for university entrance in Japan has received significant attention in the washback literature (Allen & Tahara, 2022). Recently, the introduction of four-skills exams has become a key focus (e.g., Allen, 2017; Green, 2014; Saito et al., 2022; Sato, 2019), and among the four skills, the speaking component has generated the most discussion (Allen, 2020). The crux of the issue is that while teachers are expected to follow the Course of Study guidelines for senior high schools recommended by the Ministry of Education, Culture, Sports, Science and Technology (MEXT, 2018), which recommend the teaching of four skills in five areas (reading, writing, listening, spoken production and spoken interaction), university entrance exams have been, for various reasons, largely reading-focused, and speaking ability is almost never assessed. A recent MEXT initiative to introduce external four-skills exams was intended to resolve this problem, but was ultimately rejected (Allen, 2020). Nevertheless, many universities now accept four-skills exams in their admissions systems, and various innovations have appeared regarding the assessment of speaking at different educational levels (for example Allen & Koizumi, 2024).

Given the current trend towards assessment of speaking proficiency as part of general English proficiency for university admissions, the present study focuses on a speaking test currently used for entrance to a national university in Japan. Although a traditional washback design could investigate the actual washback effects of the test in its current use, such a study is hampered by the fact that the test is taken by applicants who are distributed widely across the



country. Importantly, the purpose of this study is to look to a possible future in which the test, or a similar test of English speaking, is introduced into general university admissions in Japan. To this end, this research investigates how the test would impact the teaching of English in senior high schools if it were to be introduced as part of the Common Test, which is taken by the majority of university applicants in Japan. By identifying the predicted washback effects and mediating factors, a theory of action is presented that could support positive washback of the test use in this context

Literature review

1.1 Washback

Over 30 years of research has shown that washback is a highly complex and context-dependent phenomenon (Allen, 2025b). Attempts to explain the process have resulted in numerous models of washback, such as those by Hughes (1993), Bailey (1996), Green (2007), Shih (2009) and Sato (2019), to name but a few. Moreover, various theoretical approaches have been adopted with the aim of explaining the phenomenon, such as innovation theory (for example Wall, 2005) and sociocultural theory (Booth, 2012) (see Green, 2025, for a review). The diversity of models and theories that have been employed in washback research is testament to its scope and complexity, yet all washback models have the following three essential elements in common: a test, mediating factors and effects.

First, washback research begins with a test, which must be described according to its uses, content, task types and format, all of which contribute towards its potential for washback. Second, a range of mediating factors must be considered, all of which may affect whether or not or to what extent washback occurs. These include stakeholder- and context-related factors ranging from perceived overlap of the test with the curriculum to institutional and sociocultural norms and values of the context. Third, washback effects may be observed on teaching and learning in the form of educational content, methods and materials, and in the perceptions of stakeholders. These effects may occur in the form of changes, when some aspect of education differs across two situations (for example before and after test implementation), or reinforcement, when a test promotes the continuation of current practices, for instance those that are in accordance with curriculum aims. Both changes and reinforcements may be characterised as positive washback if they support achievement of educational goals, or negative washback if they subvert them.

The present paper targets washback as a theoretical construct, investigating the interactions between the test and the mediating factors, and the resulting effects that these interactions generate. Moreover, the intensity of the expected washback (Green, 2007; Watanabe, 2004a) and the direction of the effects (positive or negative) will be evaluated. A summary description of the mediating factors and the washback effects that will be considered, along with relevant citations from the literature, are provided in Appendix 4.

1.2 Washback research

A variety of research designs have been utilised with the aim of understanding the washback phenomenon and how it plays out in various contexts (Allen, 2025a). Researchers have utilised the before/after design to investigate the washback of a new or revised test (e.g. Andrews et al., 2002; Cheng, 2005; Wall, 2005; Wall & Alderson, 1993; Wall & Horák, 2007, 2011); the treatment/control design, which sets up a comparison of two situations, one in which the test is currently being used and one in which it is not (e.g. Alderson & Hamp-Lyons, 1996; Green, 2007; Watanabe, 1996, 1997); and other designs where the process of washback is investigated without involving a specific comparative component (e.g. Booth, 2012; Gosa, 2004; Tsagari, 2012; Tsang & Isaacs, 2022).

Washback research can contribute to the evaluation of the consequential validity of test uses. Washback studies have typically investigated the washback of tests after they have been implemented. This reflects the assumption that washback, as a consequence of test use, should be investigated following test use. In Weir's (2005) socio-cognitive framework, for example, washback research, as part of consequential validity, follows test delivery and scoring. Meanwhile, test developers and researchers have been urged to work for washback (Swain, 1985) with the understanding that engineering washback begins with test design, that is, 'washback by design' (Green, 2007). Importantly, this means not simply designing a 'good test' and hoping it will lead to positive washback but actively engaging with stakeholders and test use contexts to ensure that the conditions are in place to facilitate the intended, positive washback.

A small number of washback studies have explicitly adopted this approach (Saif, 2006; Tan & Turner, 2015). Saif (2006) conducted a needs analysis of international teaching assistants (ITAs) in an English-medium university context, developed an oral performance test based on the identified needs and investigated the washback from the newly developed test on the training of the ITAs. It was noted that the teacher's involvement in test administration and rating and her knowledge of the test were important factors mediating washback on her teaching practices. Saif's (2006) study is important because it explicitly sought to leverage test design to achieve positive washback, and it did this by involving key stakeholders (learners and teachers) in the development process.

Tan and Turner (2015) worked with teachers to align classroom-based assessment and a final writing exam. Teachers participated in rater training, using the rubric for the final exam, and this process had marked benefits on their understanding of the rubric and their ability to evaluate students' writing, which improved their ability to guide student learning during regular classes. Teachers were able to share their knowledge and experience with other teachers, extending the positive effects of the rater training sessions through the teaching community. In addition, the test developers benefited from teachers' insights and suggestions to improve the task and rubric.

Compatible with the idea of working for washback is that of the 'impact by design' approach developed by Saville (2009), explained in Saville and Khalifa (2016), and further developed and exemplified in Saville and Séguis (2025). This approach describes how washback can be

evaluated as part of test validation in practice. Consequences of test use, which include test impact and washback, should be explicitly stated and planned for as part of a theory of action (ToA). Moreover, there must be a system in place for impact research to be conducted routinely, and the system must allow developers to make changes if necessary to the test and its delivery apparatus (Saville & Khalifa, 2016: 84).

Saville's (2009) research positioned impact within a socio-cognitive model of test validity. Building on this work, the socio-cognitive model as integrated arguments approach (Chalhoub-Deville & O'Sullivan, 2020; O'Sullivan, 2025) also proposes that research into consequences is conducted at all stages in the test development cycle and incorporates the concept of 'impact-by-design' in its Theory-of-Action and Communication Arguments. These arguments address the intended and unintended effects of test use and allow test developers to create concrete plans for how to manage them. The ToA is where research into test impact is documented and where the results of such research can be incorporated to manage intended and unintended effects of test use.

One of the few examples of this approach is Green (2014), in which an impact statement and action plan were developed to support positive impact from the Test of English for Academic Purposes (TEAP). To achieve identified intended impacts, the action plan involved, for example, conducting research, publishing documentation and interacting with stakeholders through presentations. In another study, based on the stated intended consequences of test use, and an analysis of criticisms of the test in the national media, Allen and Koizumi (2024) created a ToA for a speaking test used in junior high schools in Tokyo. In both this study and Green (2014), the action plan provided concrete suggestions to test developers for supporting achievement of intended consequences.

1.3 Hypothetical washback research design

One way to investigate the potential impact of a test prior to its implementation, and develop a theory of action for its use, is to conduct a hypothetical washback study. This research design is guided by the question, what would happen if the test was introduced with a particular purpose and in a particular context? In order to answer this question, the researcher must set up a hypothetical situation in which the test is introduced as a high-stakes test. Stakeholders such as teachers and learners then imagine this situation and predict how the test would affect teaching and learning, and what mediating factors would help or hinder the generation of positive washback. The resulting data can provide insight into potential intended and unintended consequences that can feed into test development.

Numerous washback studies have employed some form of hypothetical questioning with learners and teachers (e.g. Green, 2014; Saito et al., 2022). For instance, Green (2014) investigated the proposed introduction of external four-skills exams on high-school teaching and learning through a survey completed by high-school teachers and students, within which hypothetical items were included. Saito et al. (2022) looked at the same issue of introducing four-skills tests in university admissions and conducted a meta-summary of previous



washback studies. They concluded that washback was likely to be positive, but that it would not be unequivocal nor without unintended effects. In sum, these studies demonstrate that washback researchers have already accepted the value of hypothetically framed questions that seek to estimate expected washback. Moreover, these identified studies all took place in the context of Japan in connection with the proposed use of speaking tests and four-skills tests, which suggests that such questions are ideally suited to situations where reform-based testing is being considered.

The present study

By considering the likely washback effects of exams, decision makers have more evidence at their disposal when making critical decisions about whether to introduce a specific test, or type of test, into an educational system. Such washback research is particularly important in high-stakes test use contexts, such as the assessment of English as part of university admissions.

In Japan, there are many routes into university, ranging from the general ‘written exam’ route to the interview and document-based ‘recommendation exam’ route. Many applicants also take the Common Test designed by the National Center for University Entrance Examinations, which includes an English reading paper and listening paper. Applicants prepare for the Common Test and the individual university exam(s) at school and/or by attending cram schools that provide test preparation courses (Allen, 2023).

Although the MEXT-mandated curriculum states that English should be taught and learned holistically, the current university examination system lacks a speaking assessment, leading to negative washback and the subversion of curriculum aims (e.g. Allen, 2016; Green, 2014; Watanabe, 1996, 1997). However, recent developments in technology, such as the widespread availability of tablets and laptop computers, along with the growing acceptance and use of four-skills tests of English as part of university admissions, have made speaking assessment at the national level an increasingly likely possibility. Therefore, in the aim of developing a ToA for the introduction of such assessments, the present study investigates the expected washback from a speaking assessment that was designed for use at a high-level public university in Japan.

2.1 The BCT-S

The BCT-S is a speaking test developed for admissions purposes by the British Council in collaboration with the Tokyo University of Foreign Studies (TUFS), a medium-sized national university specialising in foreign languages and humanities subjects. Since 2020, the BCT-S has been used as part of the general-route English entrance exam, which assesses all four skills with the following weightings: reading (150 points), writing (80), listening (70) and speaking (50). The primary purpose of the speaking exam is to assess speaking ability in line with the university’s admissions policy. A secondary purpose is to support MEXT curriculum goals (MEXT, 2018). The curriculum, which begins in elementary school and continues through

to senior high school, is intended to enable learners to reach the B1 level in English (MEXT, 2016).

The test design is based on Aptis General, which is a four-skills testing system developed by the British Council to assess general English (O'Sullivan et al., 2020; Shin et al., 2022). Aptis is designed for use by institutions for specific purposes (e.g. admissions, placement or exit-testing) and can be localised so that it fits the specific purpose. The BCT-S is an example of localisation at Level 2 (O'Sullivan et al., 2020: 9), which involves modifications to the Aptis General in terms of the contextual, topical and lexical content features of the test. Such modifications support alignment of test content and tasks with the Courses of Study, the goals of the university and the experiences of Japanese high-school students.

The BCT-S is a semi-direct, computer-based test in which test-takers' oral responses are recorded for scoring. The test takes approximately 12 minutes and involves four parts that include answering factual questions, describing and comparing pictures, providing opinions and a long turn (Appendix 1). Two standard setting studies were conducted linking Aptis scores to the Common European Framework of Reference for Languages (CEFR) (O'Sullivan, 2015), and the BCT-S was further defined according to the CEFR-J (Zhou et al., 2023), which employs additional levels within the more broadly defined levels (for example B1.1 and B1.2 within the B2 level) (see *BCT-S Candidate Guide*, note 1 below). Rating is conducted by trained human raters following the quality assurance system for Aptis (O'Sullivan et al., 2020: 28) and using task-based holistic scales assessing grammatical range and accuracy, lexical range and accuracy, pronunciation, fluency, and cohesion and coherence. Accommodations (i.e. large fonts, a screen reader and additional time) are provided for test-takers with special needs.

The BCT-S and Aptis are designed and validated using the socio-cognitive framework (Chalhoub-Deville & O'Sullivan, 2020; O'Sullivan & Weir, 2011; Weir, 2005). Validation of the Aptis General test is conducted internally and externally through funded research, and a series of validation studies have been conducted (see Appendix 2). However, no validation studies have yet been conducted regarding consequential validity.

2.2 Research design

This study investigates the perspectives of a core stakeholder group, that is, senior high-school teachers. Teachers play a crucial role in guiding young learners in their language study, and teachers' perspectives are paramount in the validation of reform-based tests (Winke, 2011). One difficulty in researching washback on teaching in this context, however, is that teachers and learners who perceive the BCT-S to be important are dispersed across the nation. Consequently, a traditional washback study involving classroom observations and interviews is unfeasible. On the other hand, it is possible to recruit teachers from a range of high-school contexts and to set up a hypothetical situation in which the BCT-S is perceived as a high-stakes exam. Such a research design permits the investigation of *expected* washback effects and mediating factors. Findings may contribute to the validity argument of the BCT-S in its current use, and they can be used to predict washback effects if the BCT-S is used in similar



contexts (i.e. at other universities in Japan). Moreover, this research can be used to develop a ToA for the introduction of the BCT-S, or similar speaking assessments, into the university entrance exam system in Japan. The overarching research questions were thus formulated as follows:

1. What kind of washback effects on teaching and learning in the school would senior high-school teachers in Japan expect from the BCT-S if most students were required to take it?
2. What factors would teachers expect to mediate these washback effects?

Methodology

The study adopts a qualitative research approach. Although many washback projects have utilised mixed-methods research, qualitative research has long been advocated and conducted (e.g. Booth, 2012; Sultana, 2019; Wall, 2005; Watanabe, 2004a). The study is relatively small-scale and short-term, which is appropriate for its goal, that is, to investigate stakeholder perceptions about the expected consequences of using the BCT-S that may contribute to validation efforts and ToA planning. The research is ‘a form of social inquiry’ (Hammersley, 2013: 12) that focuses on verbal data elicited from participants, each of whom represent a distinct naturally occurring case. These case studies allow the researcher to gain insight into the unique circumstances and how they interact in each situation, establishing the ‘how’ and ‘why’ (Cohen et al., 2018: 376), that is, the washback effects and the mediating factors that control them. Similarly, like most washback research, the goal is not to seek a single, generalisable outcome but to gain insight that prompts new and alternative perspectives on how washback from the BCT-S may emerge. The study is data-driven in that new interpretations of washback may be generated that are products of the unique social context of the participants; yet, at the same time, it is framed by existing washback research models and theory.

3.1 Participants

This study and all instruments were approved by the research ethics committee at the researcher’s institution. Informed consent was obtained, and teachers were paid the institutional rate for participation. *Purposive sampling* was employed, in which Japanese high-school English teachers were recruited owing to their in-depth knowledge of the context, the subject and the profession (Cohen et al., 2018). This approach is common in washback research (for example Wall & Horák, 2006). Moreover, to ensure *maximum variation* in terms of the variety of perspectives the participants brought to the study (Anderson & Arsenault, 1998), participants were sought who teach in rural and urban, public and private, and higher- and lower-ranked senior high schools in terms of academic achievement. To this end, we adopted the chain-referral method, or snowball sampling (Cohen et al., 2018). Participants were first identified from within the researcher’s network, and those participants were then asked to refer to the researcher others who, while being in comparable professional positions, were likely to hold different views owing to their contexts or experiences. Given the aims of the research, a sample size of around ten participants was targeted. Ten teachers were recruited, but, due to schedule conflicts, one was ultimately unable to participate.

Table 1: Participant information

Teacher/ School	1	2	3	4	5	6	7	8	9
Classification	Dual-route	English-focused	Vocational	Prestigious	Dual-route	Vocational	English-focused	Prestigious	National
Gender	F	M	F	F	F	F	M	M	F
Age range	45–54	45–54	45–54	55–64	18–24	55–64	35–44	55–64	45–54
No. years teaching	25	30	24	35	2	25	15	32	20
Highest degree	MA	MA	BA	MA	MA	MA	MA	MA	MA
CEFR speaking proficiency	B2–C1	C1	B2–C1	C2	B2	B2	C1	C1	C1
Study abroad months	US 24	US 12	UK 3	NZ 24	UK 10	CAN 30	UK 18	US/AUS 25	US 28
Region	Kantō	Kantō	Kantō	Shikoku	Chūbu	Shikoku	Kansai	Hokkaido	Kantō
Area	Rural, small city	Urban, large city	Rural, small city	Urban, small city	Rural, small city	Urban, small city	Urban, large city	Urban, large city	Urban, large city
Type	Public	Public	Public	Private	Public	Public	Public	Public	National
Academic level	Average	Average	Below average	High	Above average	Below average	Above average	High	Very high



Approx. no. of students	450	600	350	1600 (total) 900 in senior school	350	900	700	900	350
-------------------------	-----	-----	-----	--------------------------------------	-----	-----	-----	-----	-----

Table 1 shows that the nine teachers were generally middle-aged (seven out of the nine were 45 years or older), experienced (Mean = 23 years' teaching), highly educated (all but one had a master's degree), proficient in English (all B2 or above in speaking) and had spent some time abroad (Mean = 19 months). Teacher 5 was markedly younger and had fewer years of experience, and Teacher 3 had not spent long abroad and did not have an MA.

Table 1 also provides information about the teachers' school contexts. Taking together the survey and interview data, five groupings were created for schools that appeared to share various characteristics. *Dual-route schools*: Schools 1 and 5 are both relatively small, in rural areas and of around average academic level. Half or fewer of the students attend university, and the standard Courses of Study classes are offered (i.e. English Communication 1~3, Logic and Expression 1~3). In Grade 11, students select an academic or non-academic route, depending on whether they wish to apply to university or not. *English-focused schools*: Schools 2 and 7 are both medium-sized, in urban areas of prefectural capital cities and of around average academic level. School 2 provides additional mandatory English classes focusing on reading and presentation, while School 7 has a special English course in which English classes are conducted only in English and additional English classes and activities are provided. *Vocational schools*: Schools 3 and 6 are both vocational schools, but one is small and in a rural area, while the other is large and in an urban area. The academic level is lower, very few students go on to university and only half of the standard English classes are typically taken (i.e. English Communication 1~3). *Prestigious schools*: Schools 4 and 8 are both large, in urban centres and of high academic level. Entrance is highly competitive, and students typically aim for top-tier universities. *National school*: School 9 is a small, national school in an urban centre, with the designation of a 'Super Science High School'. It is exceptional in terms of academic level, and almost all students go to top-tier universities.

3.2 Procedure

The study had two main stages. The first stage took place prior to the interview at a time of the participant's choosing. The aim was to familiarise teachers with the test. Teachers read the *BCT-S Candidate Guide* (2021),¹ which describes the general features of the test, additional task instructions based on common errors, task-specific study advice and scoring information, including descriptions of achievement at CEFR-J bands. They also watched the *BCT-S Instructional Video*, which provides a step-by-step guide to taking the test in Japanese, and a series of four *Candidate Guidance* videos, which address the BCT-S tasks. In addition, to experience the test first-hand, participants took a timed, paper version of the test.

During this stage, participants completed an online survey (Appendix 3). The purpose of this was to provide advanced information about participants and to provide a space for them to take notes that they could refer to in the interview. The survey had two note-taking sections:

¹ https://www.tufs.ac.jp/NEWS/admission/211227_1.html; <https://www.britishcouncil.jp/exam/bct-s/test>; <https://aptisweb.com/aptis-exam-library/aptis-general>

while reviewing test information and after sitting the sample test. The questions encouraged teachers to reflect on the test and its suitability for use in their teaching context. In addition, participants answered questions about their language learning and teaching experience and about their current school. Response formats included short answer, multiple-choice and ranking. Participants spent approximately two hours on this stage.

The second stage was the interview, which lasted approximately 90 minutes and was conducted online on Zoom. The interview was semi-structured, with a range of topics covered that were guided by the research questions and by a thorough review of the washback literature (Appendix 4). It was originally anticipated that follow-up interviews may be required, but ultimately they were deemed unnecessary given the volume and detail of the interview data. The interview began by eliciting information about participants' current teaching context. This was followed by a description of the hypothetical context as follows: *I would like you to imagine that you are teaching English at your current senior high school and the BCT-S has been introduced as a high-stakes test that most of your students will need to take as part of their entrance examinations. Imagine that it has been added as part of the Common Test, for example.* Participants were then asked how, if at all, they would need to change their approach to teaching English in the school. Further questions elicited their perceptions of any mediating factors that would affect the impact of the test and any washback effects that could be expected. Particular attention was paid to the hypothetical aspect and connecting it to the reality in the classroom; that is, the researcher explicitly focused on understanding how the imagined impact would unfold in the teachers' described context of teaching. The order and inclusion of questions was not fixed; the researcher would begin by eliciting perceptions of hypothesised washback effects and then move freely between these, mediating factors, test design aspects and other topics, in accordance with the flow of the conversation. All interviews were conducted in English because participants' language proficiency was sufficient to do so.

It was emphasised to participants that the researcher was not involved in development of the BCT-S and was unaffiliated with the test developer and user. This fact promotes neutral evaluation of test washback (Winke, 2011). However, the researcher was aware that his own perceptions and beliefs about language education and assessment, if revealed during the interview, may unintentionally affect the responses of participants. Therefore, the researcher refrained from expressing agreement or disagreement but instead adopted a perspective in which he sought to understand the interviewees' perspectives. Other steps taken to demonstrate researcher integrity (Dörnyei, 2007), in addition to adopting an emic perspective and examining researcher bias, were transparency of method, thick description and contextualisation, and examination of outliers during data analysis. Each interviewee came from a unique context, which is described extensively in the narrative summaries (Appendix 5).

3.3 Data analysis

Audio data was initially transcribed using transcription software (TurboScribe, 2024). The transcription was then cross-checked against the audio and corrected as necessary. The interview data was read iteratively, and a narrative summary was drafted. This was verified by



the participant, and revisions were incorporated, leading to the final summaries (Appendix 5). The interview data was then analysed following the stages outlined for qualitative data analysis in Cohen et al (2018: 317–18). Teacher turns formed the unit of analysis, and these were coded in relation to washback effects and mediating factors (e.g. ‘washback: materials’, ‘mediator: learner interests’) and other issues (i.e. ‘suggestions for test design’). Another researcher scrutinised three of the coded transcripts to verify and query the coding, which resulted in 95.9 per cent agreement and eight instances where missing codes were added. Relationships were established between themes (i.e. effects and mediators), inferences were made and summaries were drafted. During this process, negative and discrepant cases were sought through iterative reading of the coded transcripts. Finally, a discussion of the expected washback was produced that is both grounded in the data and constructed with reference to established theories and findings (Lincoln & Guba, 1985).

Results and discussion

This section presents a concise synthesis of the interview data, culminating with a ToA to support positive washback. Extensive details and quotations are provided in the narrative summaries (Appendix 5).

4.1 Washback effects

A range of expected washback effects were elicited (Table 2) that varied according to the context and mediating factors. This section describes effects that were mentioned in three or more interviews. First, except for one teacher who believed his school was already adequately devoting time to speaking in class (7),² all of the teachers expected an increased focus on speaking activities. Three teachers (1,2,4) emphasised that students would need to learn how to structure longer monologues, which they were unaccustomed to giving.

Broadening students' experience of responding to a wide range of social topics was considered important (1,5,7,8,9). Current textbooks were believed to be overly restrictive in terms of the number of topics they covered. Moreover, students would need more experience of brainstorming and planning responses to topics under time pressure (7,8,9).

Teachers imagined how they would restructure the curriculum to accommodate teaching that would prepare students for the BCT-S. Some teachers would gradually introduce test-like tasks, thereby practising the skills required while managing the increasing difficulty of the task demands (2,3,5). In particular, more time would be spent preparing for the extended production tasks (3,7,8).

Teachers would need to become familiar with the BCT-S assessment criteria to use them to evaluate students' ability and provide appropriate feedback so that students know how to improve (2,8,9). Also, teachers would need to receive training in how to assess speaking ability (1,2,5).

Except for Teacher 7, who believed that the current materials would be sufficient, all teachers believed that at least some new materials would be required. These would typically need to include a greater range of topics and activities to develop oral skills relevant for the assessment.

² The numerical identifier is for the teacher/school to which the finding refers.

There was broad agreement that students' motivation would increase because they are typically motivated by exams (2,4,5,7,8,9) and that 'if the speaking activity follows the test, they would be motivated' (5). Teachers would also become more motivated to teach speaking in class (1,2,5), although this was more nuanced and dependent on teacher beliefs, which are discussed in the following section.

Table 2: Expected washback effects from introduction of BCT-S

	Predicted washback effect	Teacher
Speaking	More time would be spent on speaking	1,2,3,4,5,6,8,9
	planning and organising speech	1,2,4
	focusing on accuracy in speech	2,8
	pronunciation	4
	interaction activities	3
	Less time would be spent on speaking	7
Other knowledge/skills	Less time would be spent on input and comprehension	1,4
	More time would be spent on writing	1
	scriptwriting	1
	critical-thinking/higher-order skills	1
Topics	A wider range of social topics would be provided for students to consider and then respond to	1,5,7,8,9
Test-taking skills	More time would be spent on brainstorming topics under time pressure	7,8,9
Curriculum	More time would be spent on extended production activities	3,7,8
	The curriculum would be reorganised to gradually introduce students to test-like tasks	2,3,5
	An intensive preparation phase would be included in Grade 12	9
Teaching style	Teaching would need to be more 'communicative-style'	4
	Learning would need to be more individually tailored	4
	More pair work would be needed initially, followed by class presentations	8
Assessments	Teachers would need to focus on assessment criteria	2,8,9
	Teachers would need to reconcile differences in scoring criteria of BCT-S and MEXT	3,8
	Teachers may focus too much on accuracy in speaking assessments	1
	Mid-/end-of-term assessments would need to be similar to BCT-S tasks	1
Materials	New textbooks/materials would be needed	1,2,3,4,5,6,8,9
	No new textbooks/materials would be needed	7
Motivation/attitudes	Students would become more motivated to learn speaking	2,4,5,7,8,9
	Students would feel more stressed	4,8
	Teachers would become more motivated to teach speaking	1,2,5

	Teachers, students and parents would need to develop an exam-focused mentality	6
Training	Teachers would need training in how to assess speaking	1,2,5
	Teachers would need to train themselves more	2
Resources	More students would attend conversation schools	4,8
	Class sizes would need reducing	4
	More space would be needed if class sizes were reduced	4
	More teachers who are 'ideally fluent' in English would be needed	4
	Streaming students may help with classroom management	6

4.2 Washback direction, intensity and mediating factors

Almost all washback effects were considered aligned with the curriculum goals and could thus be considered positive in direction. For instance, an increased focus on speaking in classes that are intended to include speaking but currently do not could be considered positive washback. In contrast, a number of identified washback effects were considered to be potentially negative and thus worth addressing specifically in a ToA. These included the possibilities expressed by Teacher 1 that some teachers may focus on script-writing rather than spontaneous speaking, and on accuracy at the expense of fluency. Also, although including test-like tasks in classroom activities and assessments may facilitate learning if done well, if teachers focus overly on test-like tasks at the expense of other learning activities, this may be considered negative washback. This may most likely occur at schools where exams are the primary goal (i.e. 4,8). The ToA should support the integration of learning activities and test-like tasks so that classroom teaching is authentic and meaningful, rather than mechanical and focused on test-taking techniques.

The washback from introducing the test was expected to vary in intensity according to the school (Figure 1). This variation in washback intensity depends on a range of mediating factors (Table 3), which are discussed in the following sections.

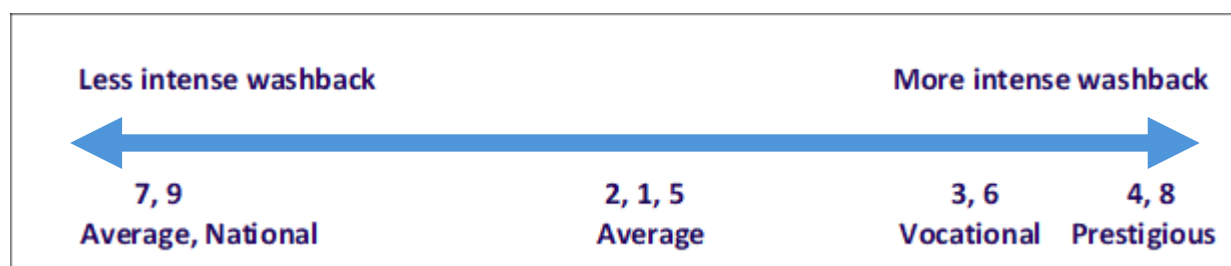


Figure 1: Washback intensity by teacher/school number and school type

Table 3: Expected mediating factors

Mediating factors
Educational goals (school mission/policy, expectations of parents and students)

Overlap between the test and current curriculum, teaching approach/methodology, materials and regular assessment practices

Proximity

Student factors (English proficiency, willingness to speak, motivation, interests)

Teacher factors (English proficiency, willingness to speak, motivation, teaching/assessment knowledge and teaching/learning beliefs)

4.3 Educational goals

A key overarching mediating factor is the type of institution and its educational mission, which influences all aspects of education, including the role and impact of exams. Whether or not students go to university and, if they do, what level of institution they target appear to be key defining characteristics. That is, the academic level of schools is crucial, with the national and prestigious schools at the highly academic end, the average/above average high schools in the middle and the vocational schools at the non-academic end. The dual-route schools share similarities with both academic and non-academic schools, according to the route being discussed.

The educational goals are clearest at the extremes. At the two vocational schools, graduates typically enter the job market and so students, teachers and parents are not exam-focused (3,6). For these students, gaining reasonable class grades is the goal. This is also the case for students who take the 'normal' or 'vocational' routes in schools that offer different courses (1,5). For English language study, this educational goal has clear consequences. Students take only three English courses, compared to the six taken at other schools, and they typically reach only a minimal level of English proficiency (i.e. CEFR A1–A2 level). Students are characterised as generally uninterested in English, which leads teachers to find ways to keep them motivated in class (for example through the use of educational technology, 3). The fact that these students do not typically take high-stakes English exams made the hypothetical situation somewhat unrealistic (this is taken up in the general discussion section). Teacher interviewees found it difficult to imagine, because introducing the BCT-S would require a fundamental change in how students, teachers and parents view exams. Moreover, the BCT-S tasks, especially Tasks 3 and 4, were perceived as excessively challenging, which could conceivably lead to negative washback.

In stark contrast, at two prestigious schools the number of students entering elite universities is the primary goal for the school and its teachers, while students aspire to, and parents demand, academic success (4,8). To be selected for top-tier universities, a higher level of English is typically necessary (i.e. CEFR B1–B2). Consequently, the teaching approach is strongly influenced by entrance exams. This is shown in the focus on skills required for these exams (i.e. reading, writing and listening) at the expense of those that are not (i.e. speaking). Similarly, the introduction of the BCT-S as a mandatory exam would be expected to have intense washback

on the education at these institutions not only because they focus on exam preparation but also because they currently omit speaking skills from the curriculum.

At the two mid-ranking public schools that emphasise English education (2,7), university exams are also considered an important goal. However, perhaps because students aim for a broader range of universities (i.e. mid- to high-ranking universities) and because of the school culture and other factors, teachers are less focused on university exams. At these schools, the development of speaking proficiency is a priority, and this is reflected in the emphasis on speaking activities and assessments. Moreover, in contrast to education at the prestigious schools, teachers use English in class and do not prioritise translation and comprehension activities. The introduction of the BCT-S would, therefore, be expected to broadly reinforce current teaching and learning practices. Teacher 7 believed the introduction of the BCT-S would have only very minimal washback because teachers are already highly focused on teaching speaking in class. Hence, the least intense washback is expected at this school. In contrast, Teacher 2 believes he would adapt his teaching considerably so that his students could get high BCT-S scores. While this demonstrates the importance of success on exams, the perceived washback effects were largely related to the specific task demands (e.g. focusing on both accuracy and fluency, doing extended monologues) rather than an increase in speaking activities.

The two dual-route schools (1,5) share characteristics with both the vocational schools (3,6) and the prestigious schools (4,8), depending on which course is being discussed. The goal of the non-academic courses is not exams, and so a wider range of activities are involved in class (1). The academic course, on the other hand, is highly exam-focused. Teachers care about the number of students gaining places at high-ranking universities and are influenced by exams, as exemplified by the inclusion of listening exam preparation courses (1,5), a focus on grammar and translation in some classes (5) and the beliefs of some teachers at the school (1,5). Consequently, washback is expected to be intense regarding the academic courses at dual-route schools, but less so for students who are not applying to university.

Finally, the national school (9) is not exam focused, despite the fact that its students typically matriculate to top-tier institutions. Perhaps owing to its designation as a 'Super Science High School', an enquiry-based approach to learning is adopted in all subjects, and teachers are expected to use innovative teaching methods. Consequently, English is typically taught in integrated-skills classes, involving both productive and receptive skills. Unlike the prestigious schools, students are expected to prepare for any entrance exam content that deviates from this curriculum in their own time. Consequently, the BCT-S is not expected to generate extensive washback at this school.

4.4 Overlap with the educational system at the school

The educational goals of a school shape the educational learning system, involving the curriculum, the delivery of that curriculum (teaching approach, methodology and materials) and assessment of achievement. Although each school is expected to follow national

guidelines that mandate the same courses, the actual content, materials and teaching approaches vary greatly. For example, although the English Communication course and its official textbooks are often characterised as reading-focused (3,5,8), speaking activities (e.g. pair work, shadowing and retelling) are incorporated in some schools (2,5,6,7,8,9), but this depends on the teacher (8) or the teacher who creates the curriculum and handouts for that grade (1,5). Consequently, the perceived alignment between this existing system and the test is a key mediating factor affecting expected washback (Alderson & Hamp-Lyons, 1996; Ferman, 2004; Green 2007; Sultana, 2019; Vyn, 2019; Wall, 2005).

The clearest example of misalignment between test and learning system is at the prestigious schools. At these schools (4,8), teaching is generally described as ‘traditional’, ‘old-fashioned’ and ‘lecture-based’. Entrance-exam-preparation ‘sub-textbooks’ are used in place of official ones (8). Classes are conducted in Japanese, except for a small number of classes that involve Assistant Language Teachers (ALTs) or native speakers of English (4,8). The teacher at one school laments that he is the only teacher who conducts his classes in English (8). Regular speaking assessments are conducted either only in specific classes (8) or not at all (4). Given these characteristics, if the BCT-S were introduced, it would generate ‘massive change’ (4). However, such extensive reform is likely to face resistance from some teachers and stakeholders (8). This issue is returned to regarding teacher beliefs.

The alignment of the test with regular speaking assessments also appears to be an important mediating factor. All schools except one (4) conduct regular end-of-term speaking assessments in classes, but this may depend on the teachers in charge of the grade (5). These performance tests come in various formats (e.g. interviews, presentations, speeches and picture description) and may involve prepared and/or spontaneous responses. Where such assessments are commonplace, teachers feel that students would already be prepared for the BCT-S (especially Tasks 1 and 2), and that some of the tasks could be adapted to more closely align with the task demands of the BCT-S. In this way, teachers expect the BCT-S to complement and expand their teaching and assessment repertoire, rather than fundamentally change it. An additional point was that students were not expected to have difficulty with the tablet-based, semi-direct format, because they are familiar with recording their voices using tablets (2,3,5,6,7,8,9).

In addition, four-skills English assessments, such as the EIKEN tests offered by the Eiken Foundation of Japan (2,8), B1 Preliminary and B2 First by Cambridge Assessment English (9) and GTEC by Benesse Corporation (5,7), are taken grade-wide in some schools, and all teachers except Teachers 1 and 3 mentioned helping students practise for speaking tests outside of classroom hours. Where such assessments are already utilised, the preparation systems in place to familiarise students with those exams were thought to be adaptable to the BCT-S. Teacher 9 suggested that she would familiarise students with the BCT-S similarly to how she does for the Cambridge Assessment exams; that is, a few months prior to the test date, she would focus on each test task in sequence, devoting one or more classes to them, in which

students would practise the tasks, analyse the criteria and give/receive feedback on their responses.

The BCT-S task demands (Appendix 1) are also an important factor mediating expected washback. Teachers typically believed that students were already being adequately prepared for Part 1 (personal Q&A) and in many cases Part 2 (picture description). These tasks were familiar to students, who engaged in similar activities in class (1,2,3,9) and in assessments (2,3), although they may still be considered very challenging for some students (2,3,4,7). In contrast, although the relevant grammar is taught (3,4,5), all interviewees believed that Part 3 (picture comparison/contrast, social Q&A) was unfamiliar to students, and that practice would be needed.

Finally, Part 4 (long turn) is considered challenging for all students and goes beyond what they are normally expected to do. For instance, while students do practise giving opinions, they have had time to digest the content through textbook activities and so are mentally prepared for it (8). The task of coming up with ideas in a limited time, then formulating and delivering a long, structured response, is beyond what most high-school students can achieve. Teachers at higher-level schools, however, believe that many of their students could cope if they had practised speaking extensively, building up through sustained classroom-based speaking as well as task-specific preparation (4,5,7,8,9).

4.5 Proximity

Teachers gave many examples of how things change in Grade 12 when the entrance exams are on the horizon, which highlights proximity as a mediating factor (e.g. Sultana, 2019; Wall & Alderson, 1993; Watanabe, 1996). Exam-focused classes and textbooks (1,5,8) become more common, and students do ‘mock tests all the time’ (8). Writing and translation instruction prepares students for entrance exams that often include such production tasks (5). Likewise, owing to the lack of speaking assessment, speaking activities and assessment become rarer (5), and in one school the sole speaking/listening class is replaced with an entrance-exam preparation class (4). Consequently, washback from the BCT-S would also likely be most intense in Grade 12, which is when washback from other exams is most obvious. However, teachers recognised that speaking should be integrated into the curriculum much earlier and cannot simply be crammed at the last minute. For example:

I think there should be more speaking activity in every class, or at least for the third-year students, because speaking ability is not something that you can build in a month or so ... you have to spend a whole year to develop your speaking skill ... yeah, there will be more time to devote to speaking activities. (9)

4.6 Learner factors

Willingness to speak is an important issue that affects teachers’ focus on speaking in class. Students are often characterised as shy (2,3) and having low self-esteem (1), so while they are ‘active towards English’, it is ‘hard work’ getting them to speak (2). In some schools, they are

willing to speak in pairs, but dislike doing so in front of whole class (3); in others, speaking is 'high pressure' because they are so unfamiliar with it (4):

... especially in my school, students are very shy to speak in front of the classroom or in front of the classmates ... they don't have confidence to speak ... Just speaking with peers is OK, but in front of the classroom, they can't say anything, too shy. So, it's a big problem for me. (3)

Some students seem to enjoy speaking activities/pair work, while others do not, which influences teacher choices about whether to do speaking activities (5,9). Class dynamics are also important. Students' willingness to speak out in class depends on not only individual but group atmosphere (9). To some extent, willingness to speak may further be influenced by students' interests and goals. In the vocational schools, although students often seem to enjoy speaking with the ALT (6) and in pairs (3), they are generally not motivated to study English (3,6). In other schools, some students are motivated by exams, and so introducing a speaking exam would likely stimulate them to speak more in class (2,4,5).

4.7 Teacher factors

In line with the washback literature (Watanabe, 2004b), teacher factors were highlighted as important mediators. These included teacher beliefs, English speaking proficiency, willingness to speak, and knowledge/experience of teaching and assessment.

Teacher beliefs not only guide an individual's behaviour but also affect teaching more broadly. Teachers emphasised that students in each grade must be taught and assessed in the same way to maintain fairness. The decision about the curriculum for each grade is the responsibility of one or two teachers, who may be rotated and may involve a senior-junior pairing. Consequently, in the aim of fairness, teachers must conform to decisions about teaching that may contradict their own preferences and beliefs. In many cases, teachers shared similar beliefs about teaching and there was little perceived conflict (2,3,4,6,7,9). However, in other cases conflict was apparent (1,5,8), and this was characterised as emerging between 'young' and 'old' teachers (1,5) and between those with experience abroad and those without (8). This conflict concerned the teaching approach, whether a communicative or traditional approach was adopted and, consequently, how speaking featured in the curriculum. One teacher (8) was transferred specifically to promote more communicative teaching practices and in particular lessons involving integrated skills and productive skills. However, teachers do not focus on speaking because it is not required in the exams:

my colleagues never practice speaking. They're always, from the start Japanese, at the end Japanese, Japanese all the time. [Students] don't talk. They always listen Japanese explanations. They're always listening. They're all doing Japanese ... so that is a problem. My previous school, all the teachers have been abroad. So, we don't have any problem at all. But here, a lot of trouble all the time. (8)

Interviewees had reasonably high English speaking proficiency (B2 or higher). However, in some schools, the other teachers are not highly proficient and would struggle with the task

demands of the BCT-S (1,4,5,6,8). These teachers would face severe difficulty teaching speaking for the test (1,4,6,8):

Not every teacher has overseas experience, and not all the teachers are fluent speakers of English, and they would have difficulty trying to produce English to the level that they would want their students to improve their speaking abilities. (4)

In some schools, teachers are generally very willing to speak in English (2,3,6,7,9). However, in others, at least some of the teachers are characterised as being very opposed to speaking English in class (1,4,5,8), and therefore there may be resistance (8).

Most teachers appeared familiar with teaching and assessing speaking. All but two teachers (3,5) have been a speaking examiner (for example for EIKEN tests). However, in schools where a traditional approach is followed, many teachers do not know how to teach speaking (4,8) and have little or no experience with speaking assessment (1,4,8). Moreover, some of the interviewees felt unsure how to use multiple criteria (8), particularly fluency and accuracy, during spontaneous speech assessment (5). Two were also unsure how they could provide individualised feedback to 40 students in a class (5,9).

4.8 Suggestions for test design

Although teachers viewed only one sample test during the familiarisation phase of this study, they were encouraged to give their opinions about the test and ask clarification questions if necessary. Consequently, teachers made a few suggestions that may be helpful to the test developer. For example, topics that are highly familiar to Japanese high-school students and involved the school context, such as institutional festivals/events, should be included (2,9); also, the picture in Part 4 could be removed because it seems redundant and potentially confusing (1). In addition, teachers were unsure about the scoring procedure explained in the instructional video (1,5,7) and whether it was human- or AI-scored (3,4,8) and double-marked or not (6). One teacher also requested additional supporting materials for teaching speaking in ways that were relevant to developing skills needed for the test (3).

4.9 Facilitating washback through a ToA

The interview data provides numerous suggestions for developing a ToA for the use of the BCT-S in the Japanese university entrance exam context (Table 4). This ToA focuses exclusively on the intended consequence of facilitating positive washback on English education and mitigating any unintended consequences that run contrary to this aim. The responsibility for carrying out the actions should be negotiated between the test developer and user.

Table 4: Theory-of-action for promoting positive consequences and mitigating negative consequences

Issue	Action
1. Teachers' English proficiency	Provide opportunities for teachers to develop their speaking ability: <i>Training sessions, subsidised English conversation courses, sustained support</i>
2. Teachers' ability to teach English speaking	Provide training for teachers in teaching speaking: <i>How to conduct meaningful and authentic in-class activities that both support students' development of speaking skills and prepare them for the test</i> <i>How to advise students about concrete 'points to focus on' that will lead to sustained improvement in speaking in general and on the test</i> <i>How to provide adequate feedback on students' spoken performance</i>
3. Teachers' ability to assess spoken English	Provide training for teachers in assessment of speaking: <i>How to use multiple criteria, such as accuracy and fluency, in assessment of students' spoken production</i>
4. Materials	Provide resources to support students' speaking development: <i>Speaking activities that cover a wide range of topics</i> <i>Activities that are communicatively authentic and meaningful while simultaneously supporting test preparation</i>
5. Teachers' understanding of the test	Provide resources that describe how the test is scored: <i>Teachers want to have concrete suggestions to give students about how to do well on the test</i>
6. Communication of counter-arguments	Provide rebuttals for arguments against the teaching and assessment of speaking: <i>Large class sizes are more conducive to lecture-style teaching and preclude teaching speaking</i> <i>Students require written English skills at university, not oral ability</i> <i>Students can start speaking later, after entering university</i> <i>Many teachers cannot speak English well enough to teach and assess speaking</i> <i>Communicative teaching results in a drop in reading and writing ability</i> <i>Speaking tests are not fairly scored</i>

7. Listening to stakeholder voices

- Make revisions to test design features and supporting materials as suggested by teachers:
Topics should be more relevant to Japanese high-school students
Provide links to resources for teaching speaking skills relevant to the assessment
-

In most schools (1,2,3,4,5,6,8), the English abilities, pedagogical expertise and assessment literacy of some teachers (Issues 1~3) were identified as inadequate. Because these inadequacies may hinder positive washback, they would need to be addressed. Such issues were particularly prominent at schools where teaching and assessment of speaking were largely absent (4,8). Therefore, the ToA needs to assume diversification according to school context, such that some schools will require more targeted intervention than others. For example, schools that are identified as having poor resources for teaching speaking should be allocated additional support in terms of the number of ALTs and training sessions provided. Importantly, concerns about teacher factors were also identified among high-school English teachers in South Korea when faced with the introduction of the National English Ability Test (NEAT) (Whitehead, 2016). Given that the NEAT was ultimately discontinued, resulting in significant expenditure of public funding (Shin, 2024), such concerns must be addressed in a ToA.

The suggestions to improve teachers' knowledge and skills may also prevent occurrence of unintended consequences. Interviewees raised concerns that teachers who were unfamiliar with teaching and assessing speaking may introduce practices that could negatively impact learning, such as overly focusing on accuracy in speaking assessment and on speech writing at the expense of spontaneous production activities (1). Also, teachers may feel inclined to 'coach' students to get high scores (2). Therefore, advising them how to help learners to develop their speaking abilities while not over-relying on test-taking strategies is key to avoiding negative washback. Achieving this will require effective communication with teachers in addition to training sessions and materials (Issue 4).

There were also indications that teachers may not sufficiently understand the information presented about how spoken responses were scored (Issue 5). One teacher considered the explanation of score assignment too 'general' and therefore neither students nor teachers would know how to improve their performance (2). Other teachers had questions about the scoring procedures, suggesting that the test documentation may need revising or supplementing. Another possibility is to provide training that involves selected teachers in the rating process so that they better understand how it functions and how they can integrate it with their own teaching and assessment practices (e.g. Saif, 2006; Tan & Turner, 2015). Positive washback may be further generated through a trickle-down effect where teachers share their experience with other teachers (Vyn, 2019).



One of the interviewees (4) raised a number of counter-arguments to teaching and assessing speaking, thereby supporting the status quo in the school (Issue 6). Such arguments are common among stakeholders who are resistant to teaching and assessment innovations (Allen, 2020). Any serious attempt to introduce speaking assessments into the Japanese entrance exam system must therefore produce a comprehensive ToA to support change at these institutions. Previous innovations in English speaking assessments have faced considerable resistance, although many public criticisms may be unsubstantiated or inaccurate (Allen & Koizumi, 2024). Because such criticisms may mislead stakeholders and ultimately impede positive washback, a ToA should include research that addresses such criticisms and provides rebuttals where needed. Moreover, test developers/users must effectively communicate with stakeholders regarding these issues. This highlights the necessity of an effective communication plan that can assist in the implementation of a ToA (Allen & Koizumi, 2024; O'Sullivan, 2025).

Finally, the interviews provided a number of suggestions on test design that may lead to revisions of the test tasks and presentation (Issue 7). Listening to stakeholder voices and involving them in the process of test design in this way may further support the goal of working for washback. Additionally, the BCT-S was localised with input from students and staff at a specific institution to make the test appropriate for the background and level of the test-takers there. However, if the BCT-S were to be used across various institutions, the test's suitability for the new broader context and stakeholders would need to be considered.

Limitations and future directions

A limitation of the present study was in the application of the hypothetical situation in schools where learners typically do not take high-stakes exams, that is, at the vocational schools (3,6) in particular. Teachers struggled to imagine the hypothetical introduction of the BCT-S, and therefore the inferences that can be drawn from those interviews were limited. Although the information was still useful for understanding the range of educational contexts and how they may or may not be affected by the introduction of speaking assessments, for hypothetical washback designs to produce findings that are of use for a ToA, it is imperative that the hypothesised situation is perceived as realistic by participants.

Also, the present study involved eight public/national schools and only one private school. Because roughly one in four senior high schools are private, it would have been helpful to have a greater sample from this sector. Furthermore, teachers opposed to teaching and assessing speaking were regularly described by interviewees, but they were not represented among interviewees. Although it is challenging to recruit participants who oppose the teaching and assessment of speaking, such detracting viewpoints are also important for developing a ToA for testing innovations and should ideally be included in future studies.

In addition, to further refine the proposed ToA, it would be worthwhile to communicate with the same and/or different teachers, and other stakeholders, regarding the issues covered and the actions proposed. If the BCT-S or similar test were to be introduced on a large scale in the present context, a follow-up 'actual' washback study would also be complementary. Finally, further research into development and implementation of ToAs as part of working for washback is essential. It is hoped that such endeavours will support the generation of tangible positive consequences for language learners and relevant stakeholders.



Acknowledgements

This research was funded by a British Council Assessment Research Grant (2023–03). Thank you to Dr Yujia Zhou for providing background information on the BCT-S (Appendix 2) and Prof. Tony Green for feedback on earlier drafts of this article.

References

- Alderson, J.C. & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13, 280–97. <https://doi.org/10.1177/026553229601300304>.
- Allen, D. (2016). Japanese cram schools and entrance exam washback. *Asian Journal of Applied Linguistics*, 3(1), 54–67. <https://caes.hku.hk/qjal/index.php/qjal/article/view/338>
- Allen, D. (2017). Investigating Japanese university students' English language proficiency: Predicting factors and pre-tertiary washback. *IELTS Partnership Research Papers*, 2. IELTS Partners. <https://www.ielts.org/teaching-and-research/research-reports>
- Allen, D. (2020). Proposing change in university entrance examinations: A tale of two metaphors. *Shiken*, 24(2), 23–38. https://teval.jalt.org/sites/default/files/24_02_23_Allen_Two_Proposals.pdf
- Allen, D. (2025a). Washback research designs. In D. Allen (ed.) *Washback research in language assessment: Fundamentals and contexts* (pp. 52–64). Routledge.
- Allen, D. (2025b). *Washback research in language assessment: Fundamentals and contexts*. Routledge.
- Allen, D. & Koizumi, R. (2024). The *English Speaking Achievement Test for Japanese Junior High School Students* (ESAT-J): An analysis of public criticisms using the sociocognitive framework. *Language Assessment Quarterly*, 4–5, 473–505. <https://doi.org/10.1080/15434303.2024.2438982>
- Allen, D. & Tahara, T. (2022). A review of washback research in Japan. *JLTA Journal*, 24, 3–22. https://doi.org/10.20622/jltajournal.24.0_3
- Anderson, G. & Arsenault, N. (1998). *Fundamentals of educational research* (2nd edn). Routledge. <https://doi.org/10.4324/9780203978221>
- Andrews, S.J., Fullilove, J. & Wong, Y. (2002). Targeting washback: A case study. *System*, 30, 207–233.
- Bachman, L.F. & Palmer, A.S. (2010). *Language assessment in practice*. Oxford University Press.
- Bailey, K.M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13, 257–79. <https://doi.org/10.1177/026553229601300303>.
- Booth, D.K. (2012). *Exploring the washback of the TOEIC in South Korea: A sociocultural perspective on student test activity* [Doctoral dissertation, University of Auckland]. University of Auckland Libraries and Learning Services. <http://hdl.handle.net/2292/19379>
- Chalhoub-Deville, M. & O'Sullivan, B. (2020). *Validity: Theoretical development and integrated arguments*. British Council Monograph Series 3. Equinox Publishing.

- Chapman, D.W. & Snyder, C.W. (2000). Can high stakes national testing improve instruction: Re-examining conventional wisdom. *International Journal of Educational Development*, 20, 457–474. [https://doi.org/10.1016/S0738-0593\(00\)00020-1](https://doi.org/10.1016/S0738-0593(00)00020-1)
- Cheng, L. (2005). *Changing language teaching through language testing*. Cambridge University Press.
- Cohen, L., Manion, L. & Morrison, K. (2018). *Research methods in education* (8th edn). Routledge.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press.
- Ferman, I. (2004). The washback of an EFL national oral matriculation test to teaching and learning. In L. Cheng & Y. Watanabe (eds) with A. Curtis, *Washback in language testing: Research contexts and methods* (pp. 190–210). Lawrence Erlbaum.
- Green, A. (2007). *IELTS washback in context: Preparation for academic writing in higher education*. Studies in Language Testing 25. UCLES/Cambridge University Press.
- Green, A. (2014). *The Test of English for Academic Purposes (TEAP) impact study: Report 1 – Preliminary questionnaires to Japanese high school learners and teachers*. Eiken Foundation of Japan. https://www.eiken.or.jp/teap/group/pdf/teap_washback_study.pdf
- Green, A. (2025). Theories and models of washback. In D. Allen (ed.) *Washback research in language assessment: Fundamentals and contexts* (pp. 26–39). Routledge.
- Gosa, C.M.C. (2004). *Investigating washback: A case study using student diaries*. [Unpublished doctoral dissertation]. Lancaster University.
- Hammersley, M. (2013). *What is qualitative research?* Bloomsbury Academic.
- Hughes, A. (1993). *Backwash and TOEFL 2000*. [Unpublished manuscript]. University of Reading.
- Hughes, A. (2003). *Testing for language teachers*. Cambridge University Press.
- Kane, M.T. (2013). Validating the interpretation and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73. <https://doi.org/10.1111/jedm.12000>
- Lincoln, Y.S. & Guba, E.G. (1985). *Naturalistic enquiry*. Sage.
- MEXT (Ministry of Education, Culture, Sports, Science and Technology) (2016). *Gurōbaru-ka ni taiō shita eigokyoiku kaikaku jisshi keikaku* [Implementation plan for English education reform in response to globalisation.] https://www.mext.go.jp/a_menu/kokusai/gaikokugo/_icsFiles/afieldfile/2014/01/31/1343704_01.pdf
- MEXT (2018). *Gaikoku-go-hen eigo-hen: Kōtōgakkō gakushū shidō yōryō (Heisei 30-nen kokuji) kaisetsu* [Foreign languages (English): Explanation of high school curriculum guidelines (2018 notification)]. https://www.mext.go.jp/content/1407073_09_1_2.pdf
- O’Sullivan, B. (2015). *Linking the Aptis reporting scales to the CEFR*. TR/2015/003. British Council. www.britishcouncil.org/sites/default/files/tech_003_barry_osullivan_linking_aptis_v4_single_pages_0.pdf
- O’Sullivan, B. (2025). Washback in test validation. In D. Allen (ed.) *Washback research in language assessment: Fundamentals and contexts* (pp. 40–51). Routledge.

- O'Sullivan, B. & Weir, C.J. (2011). Test development and validation. In B. O'Sullivan (ed.) *Language testing: Theories and practices* (pp. 13–32). Palgrave.
- O'Sullivan, B., Dunlea, J., Spiby, R., Westbrook, C. & Dunn, K. (2020). Aptis General technical manual. Version 2.2. *Aptis Technical Report. TR/2020/001*. British Council.
https://www.britishcouncil.org/sites/default/files/aptis_technical_manual_v_2.2_final.pdf
- Saif, S. (2006). Aiming for positive washback: A case study of international teaching assistants. *Language Testing*, 23(1) 1–34. <https://doi.org/10.1191/0265532206lt322oa>
- Saito, H., Sawaki, Y. & Kasahara, K. (2022). Evaluating fairness and justice of external English language test score interpretation and use for Japanese university admission. *Language Assessment Quarterly*, 19(4), 422–448.
<https://doi.org/10.1080/15434303.2022.2083965>
- Sato, T. (2019). An investigation of factors involved in Japanese students' English learning behavior during test preparation. *Papers in Language Testing and Assessment*, 8(1), 69–95. <https://doi.org/10.58379/FSBQ6351>
- Saville, N. & Khalifa, H. (2016). The impact of language assessment. In D. Tsagari & J. Banerjee (eds), *Handbook of second language assessment* (pp. 77–96). De Gruyter.
- Saville, N., & Séguis, B. (2025). Washback, impact, and consequences. In D. Allen (ed.) *Washback research in language assessment: Fundamentals and contexts* (pp. 12–25). Routledge.
- Shih, C.M. (2009). How tests change teaching: A model for reference. *English Teaching: Practice and Critique*, 8(2), 188–206. <https://files.eric.ed.gov/fulltext/EJ859684.pdf>
- Shin, D. (2024). Critical discursive approaches to evaluating policy-driven testing: Social impact as a target for validation. *Language Testing*, 41(1), 162–180.
<https://doi.org/10.1177/02655322231163863>
- Shin, J.-Y., Rodríguez-Fuentes, R.A., Swatek, A.M. & Ginther, A. (2022). Aptis test review. *Language Testing*, 39(1), 172–187. <https://doi.org/10.1177/02655322211032873>
- Sultana, N. (2019). *Exploring the alignment of the Secondary School Certificate English Examination with curriculum and classroom instruction: A washback study in Bangladesh* [Doctoral dissertation, Queen's University]. Queen's Graduate Theses and Dissertations.
<https://qspace.library.queensu.ca/items/29801b52-8f27-4b89-94b3-0904aa35e673>
- Swain, M. (1985). Large-scale communicative language testing: A case study. In Y.P. Lee, C.Y.Y. Fok, R. Lord & G. Low (eds), *New directions in language testing*. Pergamon Press.
- Tan, M. & Turner, C.E. (2015). The impact of communication and collaboration between test developers and teachers on a high-stakes ESL exam: Aligning external assessment and classroom practices. *Language Assessment Quarterly*, 12(1), 29–49.
<https://doi.org/10.1080/15434303.2014.1003301>
- Tsagari, D. (2012). FCE exam preparation discourses: Insights from an ethnographic study. *Cambridge ESOL: Research Notes*, 47, 36–48.
<https://www.cambridgeenglish.org/Images/22669-rv-research-notes-47.pdf>

- Tsang, C.L. & Isaacs, T. (2022). Hong Kong secondary students' perspectives on selecting test difficulty level and learner washback: Effects of a graded approach to assessment. *Language Testing*, 39(2), 212–238. <https://doi.org/10.1177/02655322211050600>
- Turboscribe. (2024). [Computer software]. <https://turboscribe.ai>
- Vyn, R. (2019). *Promoting curricular innovation through language performance assessment: Leveraging AAPPL washback in a K–12 world languages program*. [Doctoral dissertation, University of Iowa]. Iowa Research Online. <https://doi.org/10.17077/etd.iepe-7s2s>
- Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory*. Cambridge University Press.
- Wall, D. & Alderson, J.C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*, 10(1), 41–69. <https://doi.org/10.1177/026553229301000103>
- Wall, D. & Horák, T. (2006). *The impact of changes in the TOEFL examination on teaching and learning in Central and Eastern Europe – Phase 1: The baseline study*. Educational Testing Service. <https://doi.org/10.1002/j.2333-8504.2006.tb02024.x>
- Wall, D. & Horák, T. (2007). Using baseline studies in the investigation of test impact. *Assessment in Education: Principles, Policy & Practice*, 14(1), 99–116. <http://dx.doi.org/10.1080/09695940701272922>
- Wall, D. & Horák, T. (2011). *The impact of changes in the TOEFL® exam on teaching in a sample of countries in Europe: Phase 3, The role of the coursebook; Phase 4, Describing change*. Educational Testing Service. <https://doi.org/10.1002/j.2333-8504.2011.tb02277.x>
- Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing*, 13(3), 318–333. <https://doi.org/10.1177/026553229601300306>
- Watanabe, Y. (1997). *The washback effects of the Japanese university entrance examinations of English-classroom-based research*. [Unpublished doctoral dissertation]. Lancaster University.
- Watanabe, Y. (2004a). Methodology in washback studies. In L. Cheng & Y. Watanabe (eds) with A. Curtis, *Washback in language testing: Research contexts and methods*, (pp. 19–36). Lawrence Erlbaum.
- Watanabe, Y. (2004b). Teacher factors mediating washback. In L. Cheng & Y. Watanabe (eds) with A. Curtis, *Washback in language testing: Research contexts and methods*, (pp. 129–146). Lawrence Erlbaum.
- Weir, C.J. (2005). *Language testing and validation: An evidence-based approach*. Palgrave-Macmillan.
- Whitehead, G.E.K. (2016). The rise and fall of the National English Ability Test: Exploring the perspectives of Korean high school English teachers. *Asian EFL Journal*, 8(4), 124–155. <https://www.asian-efl-journal.com/main-editions-new/volume-18-issue-4-december-2016-quarterly-journal/index.htm>
- Winke, P. (2011). Evaluating the validity of a high-stakes ESL test: Why teachers' perceptions matter. *TESOL Quarterly*, 45(4), 628–660. <https://doi.org/10.5054/tq.2011.268063>



Zhou, Y.J., Motteram, J. & Negishi, M. (2023, September). *Exploring the link between BCT-S and the CEFR-J: Focusing on candidates' self-assessment*. Paper presented at the 9th Annual International Conference of the Asian Association for Language Assessment, Chuo University, Tokyo, Japan.

Appendices

Appendix 1: BCT-S test design

Part	Task	Preparation time	Expected response length	CEFR-J range assessed	Skills focus	Rating criteria
1	Three questions on familiar topics	None	30 seconds (each question)	A0–A2.2	Giving personal information	Separate task-based holistic scales for each part
2	One question about a picture (for example fishing), followed by two questions concerning the topic associated with the picture (for example hobbies)	None	45 seconds (each question)	A0–B1.2	Describing, expressing opinions, providing reasons and explanations	Aspects of performance assessed include:
3	Describe, compare and contrast two pictures (for example showing different forms of communication) and provide an opinion for two questions on this topic supported by reasoning	None	45 seconds (each question)	A0–B1.2	Describing, comparing and contrasting, providing reasons and explanations	1) grammatical range and accuracy 2) lexical range and accuracy 3) pronunciation
4	Long turn on a more abstract topic with three prompts (for example overcoming challenges)	1 minute	2 minutes (entire response)	A1/A2–C1	Integrating ideas on an abstract topic into a long turn. Giving and justifying opinions, advantages and disadvantages	4) fluency 5) cohesion and coherence

Reference: *BCT-S Candidate Guide*; *BCT-S Instructional Videos*; *Aptis General Guide* (O'Sullivan et al., 2020: 22)

www.britishcouncil.org/english-assessment/english-language-research

This report is brought to you by English Language Research, British Council

Appendix 2: BCT-S-related research presentations

- [12] Zhou, Y.J., Negishi, M., Yoshitomi, A., Kurokawa, S. & Carpenter, J. (2023b, September). Exploring Target Language Use tasks for the BCT-S: A content analysis of English course syllabi. Paper presented at the *26th Japan Language Testing Association (JLTA) Annual Conference*, Tohoku University, Japan.
- [11] Zhou, Y.J., Motteram, J. & Negishi, M. (2023a, September). Exploring the link between BCT-S and the CEFR-J: Focusing on candidates' self-assessment. Paper presented at the *9th Annual International Conference of the Asian Association for Language Assessment*, Chuo University, Tokyo, Japan.
- [10] Zhou, Y.J., Motteram, J. & Negishi, M. (2022, December). Localisation and test validation: The case of BCT-S. Paper presented at the *10th British Council New Directions in English Language Assessment Conference*, Bali, Indonesia.
- [09] Zhou, Y.J., Negishi, M. & Yoshitomi, A. (2022, November). Investigating the appropriateness of using a speaking test for university admissions as a placement test. Paper presented at the *25th Japan Language Testing Association (JLTA) Annual Conference*, online.
- [08] Zhou, Y.J., Motteram, J., Sudheendra, S. & Yasuda, C. (2021, December). Tablet-delivered speaking tests for Japan's university entrance exams. Paper presented at the *9th British Council New Directions in English Language Assessment Conference*, online.
- [07] Zhou, Y.J., O'Sullivan, B., Negishi, M. & Yoshitomi, A. (2021, September). Exploring the effectiveness of planning time in a computer-delivered speaking test. Paper presented at the *24th Japan Language Testing Association (JLTA) Annual Conference*, online.
- [06] Motteram, J., Zhou, Y.J. & Negishi, M. (2021, August). Linking a computer-based test of speaking to the CEFR-J: Looking for evidence of can-dos. *The JACET 60th Commemorative International Convention*, online.
- [05] Zhou, Y.J., Rutherford, K., Fairbairn, J., Negishi, M., Dunlea, J. & O'Sullivan, B. (2019, December). Language assessment transition in a Japanese university. Paper presented at the *7th British Council New Directions in English Language Assessment Conference*, Yokohama, Japan.
- [04] Zhou, Y.J., Negishi, M. & Yoshitomi, A. (2019, November). High school students' perceptions of a computer-based speaking test for Japanese university admission. Paper presented at the *2nd Japan Association for Applied Linguistics (JAAL) Conference in JACET*, Takachiho University, Tokyo.
- [03] Zhou, Y.J., Dunlea, J., Negishi, M., O'Sullivan, B. & Yoshitomi, A. (2019, September). Gathering a posteriori validity evidence of a computer-based speaking test for Japanese university admission. Paper presented at the *22nd Japan Language Testing Association Annual Conference (JLTA)*, Niigata Seiryō University, Niigata Prefecture.
- [02] Zhou, Y.J., Dunlea, J., Negishi, M. & Yoshitomi, A. (2018, December). Collecting a priori validity evidence during the development of a computer-based speaking test for Japanese university admission purposes. Paper presented at the *1st JAAL (The Japan Association for Applied Linguistics Conference in JACET)*, Takachiho University, Tokyo.

- 
- [01] Zhou, Y.J., Dunlea, J., Negishi, M. & Yoshitomi, A. (2018, October). Localisation of an international speaking test for Japanese university admission. Paper presented at the *6th British Council New Directions in English Language Assessment Conference*, Kuala Lumpur, Malaysia.

Appendix 3: Survey content

Section 1: While reviewing the test information

While reviewing the test information (BCT-S Candidate Guide, the five short videos about the test tasks and scoring), feel free to note any thoughts that come to mind. The notes do not need to be in full sentences. They do not need to be explained. The notes are to be used to help you remember when we discuss the test during the interview. (Your notes can be written in English and/or Japanese, whichever is convenient.)

1. Do you think this test covers the types of English language speaking knowledge and abilities that are taught in senior high school English classes?
Is there anything you feel is not assessed?
Is there anything assessed that is not covered in classes?
2. Do you have any questions about the test?
Do you want to know more about any aspect of how the test is developed, or how it is assessed?
Feel free to make a note of any questions here.

Section 2: After sitting the sample test

After taking the BCT-S sample test, feel free to note any thoughts that come to mind. The notes do not need to be in full sentences. They do not need to be explained. The notes are to be used to help you remember when we discuss the test during the interview. (Your notes can be written in English and/or Japanese, whichever is convenient.)

1. What are your initial impressions of the test?
What do you like or not like about it?
What was easy or difficult about it?
What did you think about each task?
2. If your students had to take this test, how well do you think they could cope with the test tasks?
What would they be able to do well?
What may they have difficulty with?

Section 3: Background information

Please provide some basic information about yourself and your English learning and teaching experiences. Your answers should be brief. If necessary we can follow up on any important points in the interview.

1. How old are you? 18-24 25-34 35-44 55-64 65+
2. What is your gender? Female Male Prefer not to say
3. What subject was your degree in? (e.g., sociology)
4. If you hold a higher degree (e.g., MA), what was the subject of the degree?
5. What teaching degree qualifications do you hold? (e.g., Japanese teaching licence for Junior High School, CELTA).

6. Please estimate your overall English language ability according to the CEFR (A2~C2).
(If you are unsure about the CEFR, skip this question).
A2 B1 B2 C1 C2
7. Please estimate your speaking ability in English according to the CEFR. (If you are unsure about the CEFR, skip this question).
A2 B1 B2 C1 C2
8. If you have taken an English test in the past, write the name, score and date (e.g., EIKEN, Level Pre-1, around 2010).
9. If you have taken an English Speaking test in the past, write the name, score and date (e.g., IELTS Speaking, Band 5.5, around 2020).
10. Which of the 4 skills do you feel most confident in? Please rank in order from most (top) to least confident (bottom).
Reading Writing Listening Speaking
11. Have you lived and/or studied in an English-speaking country for before? If so, where and for how long? (e.g., Australia, 4 months)
12. Do you have any other significant English learning experiences? (e.g., English-speaking family members)
13. How many years have you been a teacher?
14. How many schools have you taught at prior to your current school?
15. At which levels of education do you have experience of teaching English? Please check all that apply.
Elementary school Junior high school Senior high school Higher education
Supplementary (cram/conversation school)
16. Which kinds of high schools have you taught in? Please select all that apply.
Public (i.e., local, prefectural, national) Private Vocational Rural
Suburban Urban Lower-ranked (in terms of academic achievement)
Middle-ranked Higher-ranked
17. How much experience do you have of making English tests?
A lot A moderate amount A little None at all
18. How much experience do you have with making and/or administering speaking tests?
A lot A moderate amount A little None at all
19. How familiar are you with English speaking tests? (e.g., EIKEN)
Very familiar Somewhat familiar Not so familiar Not at all familiar
20. Have you ever trained as a speaking examiner for a large-scale exam? (e.g., EIKEN)

Yes No

Section 4: Your current school

Please provide some information about your current school. In this research, we will make sure that you or your school cannot be identified.

21. What prefecture is your current school in?

22. What kind of school is it? Please select all that apply.

Public (i.e., local, prefectural, national) Private Vocational Rural
Suburban Urban Lower-ranked (in terms of academic achievement)
Middle-ranked Higher-ranked Junior high Senior high Comprehensive
(junior and senior combined) Co-educational Single-sex (boys)
Single-sex (girls)

23. Is there any other characteristic that makes your school different from other schools?
(e.g., it prioritizes sports, it is a religious school)

Appendix 4: Interview topics, supporting research, and example questions

1. The hypothetical situation

I would like you to imagine that you are teaching English at your current senior high school and the BCT-S has been introduced as a high-stakes test that most of your students will need to take as part of their entrance examinations. Imagine that it has been added as part of the Common Test, for example.

2. Questions relating to washback effects from the test

Questions regarding washback effects were developed considering the research questions and the washback literature.

Washback Effects	Research has shown that ...	Questions
Content Methods Materials Perceptions Motivation Feelings	Tests can influence teaching content and/or methods (Abdulhamid, 2018; Ahmed, 2018; Alderson & Hamp-Lyons, 1996; Cheng, 1997, 2005; Ferman, 2004; Shih, 2009; Shohamy et al., 1996; Sultana, 2019; Tsagari, 2012; Umashankar, 2017; Vyn, 2019; Wall & Alderson, 1993; Wall & Horák, 2011; Watanabe, 1996, 1997), including classroom assessments (Ahmed, 2018; Shih, 2009), and choice of materials (Allen et al., 2024; Cheng, 2005; Shohamy et al., 1993; Shohamy et al., 1996; Sultana, 2019). Teachers' perceptions and motivation towards teaching may change due to test use (Cheng, 2005; Shohamy 1993). The use of tests can impact teachers' feelings, for example creating stress and anxiety (e.g., Abdulhamid, 2018; Ferman, 2004; Shohamy et al., 1996).	• How would you change your teaching, if at all? Why? • Would you do more or less of anything? (i.e., time spent) Why? • Would you change the content of your English classes? • Would you change your methods of teaching English in class? • Would you change the materials of your English classes? • Would your perceptions about teaching speaking change? (e.g., importance) • Would your motivation to teach speaking change? • Would your feelings about teaching English, or speaking in particular, change? (e.g., anxiety, stress, enjoyment)

3. Mediating factors of washback from the speaking assessment and suggested interview questions

Questions regarding mediating factors were derived from the washback literature.

Mediating Factor	Research has shown that..	Sample questions
Test format	Teachers' perceptions of learner familiarity with the test format may influence classroom instruction (Allen et al, 2024). The test format can influence teaching content (Wall & Horák, 2011).	• Are your students familiar with the format?
Task demands	Teachers' perceptions and understanding of task demands can influence teaching (Green, 2007; Qi, 2004; Vyn, 2019; Wall, 2005; Wall & Horák, 2011; Watanabe, 2004).	• What do you think about the four BCT-S tasks?
Alignment with curriculum/ standards	The perceived overlap of the test with the target construct is believed to determine washback potential (Green, 2007). Teachers' perceptions about the coverage and relevance of the test tasks in relation to the curriculum can influence their teaching (Allen et al, 2024; Ahmed, 2018; Burrows, 2004; Shohamy et al, 1996; Sultana, 2019; Vyn, 2019; Wall, 2005). For example, Ahmed (2018) found that although writing was intended to be taught in the school curriculum, teachers did not teach writing because it was not included in a high-stakes exam.	• How compatible are these tasks with the content and aims of the Courses of Study for English? • Is there any aspect of speaking that you feel is not assessed but should be? • Is there anything that is irrelevant to classroom teaching / speaking ability?
Alignment with teaching	Teachers' perceptions about the compatibility of the test tasks with their current teaching can influence their teaching (e.g., Alderson & Hamp-Lyons, 1996; Allen et al, 2024; Vyn, 2019).	• How compatible are these tasks with your teaching?
Alignment with materials	Teachers' perceptions about the compatibility of the test tasks with their current teaching materials can mediate washback effects (Ahmed, 2018; Alderson & Hamp-Lyons, 1996; Wall, 2005; Wall & Horák, 2011).	• Are the materials you have adequate to help students develop the speaking skills for the test?

Scoring criteria	Teachers' perceptions of the scoring criteria and rubric can influence their teaching (Tan & Turner, 2015).	• What do you think about the scoring criteria? • Are they appropriate?
Student resources	The time that test takers have to devote to learning (Allen, 2017; Wall & Horák, 2011), as well as their motivation (Alderson & Hamp-Lyons, 1996; Allen, 2017; Wall, 2005) and ability to learn (Allen, 2017; Wall, 2005) can mediate washback effects.	• Do the students have the time, ability, and motivation to study speaking?
Perceived difficulty	Perceived test difficulty in relation to learner proficiency can influence teaching (Cheng, 2005; Ferman, 2004; Green, 2006, 2007; Qi, 2004; Shih, 2009; Wang, 2010; Wall, 2005; Watanabe, 1996, 2004).	• Is the level of difficulty suitable for the range of proficiencies in your classes?
Perceived importance	Perceived importance can influence teaching (e.g., Abdulhamid, 2018; Cheng, 2005; Green, 2006, 2007; Shih, 2009; Sultana, 2019; Yu, 2010) and learning (Allen, 2017).	• Not applicable. This is controlled by in the hypothetical situation.
Understanding of the exam	Teachers' understanding and access to information about the test may influence teaching (Sultana, 2019; Wall, 2005; Wall & Horák, 2011).	• Do you have any questions about the exam? Is there anything you do not understand about the exam?
Teacher beliefs about teaching and assessment	Teachers' beliefs about teaching (Abdulhamid, 2018; Ahmed, 2018; Shih, 2009; Sultana, 2019; Tsagari, 2012; Wall, 2005; Wang, 2010; Watanabe, 1996, 1997) and assessment (Ahmed, 2018; Tsagari, 2012; Wall, 2005; Yu, 2010) can mediate washback.	• Do your students need to learn speaking skills? • Do you have experience of introducing any speaking tests into your school/class? Can you describe them? Were they successful? Why (not)?

Teacher beliefs about test preparation	Teachers' beliefs about optimum ways to prepare for tests can influence their instruction (Shih, 2009; Sultana, 2019; Tsagari, 2012; Wall & Horák, 2011; Watanabe, 1996, 1997).	• What are your views about test preparation?
Teacher training and experience, knowledge of and confidence in teaching	Teachers' experience, education and training (Sultana, 2019; Tsagari, 2012; Vyn, 2019; Wall, 2005; Wall & Horák, 2011; Watanabe, 1996, 1997, 2004; Yu, 2010), and their knowledge of teaching practices (Chapman & Snyder, 2000; Wall, 2005; Wall & Alderson, 1993; Wang, 2010; Yu, 2010) can influence classroom instruction. Teachers' own language proficiency can influence their approach (Wall & Horák, 2011).	• Do you believe you are sufficiently trained to teach speaking for this test? • How confident are you about teaching speaking in class?
Teaching resources	The availability of resources, such as the time available for teaching (Abdulhamid, 2018; Ahmed, 2018; Wall, 2005; Yu, 2010), the number of available teachers (Wall, 2005), class sizes and physical environment (Umashankar, 2017; Wall, 2005; Wang, 2010), classroom equipment and materials (Allen et al., 2024; Wall, 2005; Wall & Alderson, 1993; Wall & Horák, 2011; Yu, 2010) can all influence teaching.	• Do you have the necessary time, classroom equipment, human resources and anything else to teach speaking in class?
Extrinsic mediating factors - Proximity of exam - Institutional factors - Political factors - Geographical factors - Cultural factors	Tests may generate stronger washback effects in the immediate run up to the test (e.g., Shohamy et al., 1996; Sultana, 2019; Wall & Alderson, 1993; Watanabe, 1996). Institutional factors including nation-wide policies (Ahmed, 2018) and the school culture, policy and management priorities (Shih, 2009; Wall & Horák, 2011; Watanabe, 2004; Yu, 2010) can mediate washback. Local political factors may influence teacher's approaches to teaching and assessment (e.g., Ahmed, 2018; Umashankar, 2017; Wall, 2005). Geographical factors may affect the time required to travel to school and also the school conditions, which can mediate test washback (e.g., Umashankar, 2017; Wall, 2005). Cultural factors, such as the traditional approaches to teaching and	• Is there anything that hinders you, or prevents you from, teaching English speaking as recommended in the Courses of Study? • If so, would this factor also hinder you in teaching English speaking for the test? • Do you think other exams would prevent you from focusing on speaking?

- Prevalence of private tutoring	assessment, may mediate test washback (Yu, 2010). The availability and use of private tutoring to support out-of-class learning can affect teaching (e.g., Allen, 2016, 2023; Wall, 2005) and learning (Tsang & Isaacs, 2022).	• Do you think that any out-of-school education, like cram schools, would have any impact on teaching speaking in school?
- Parental involvement and expectations	The degree to which parents are involved in education and their expectations can influence teaching (Ahmed, 2018) and learning (Ferman, 2004; Tsang & Isaacs, 2022).	• How about the impact of parents, or school principals, or other stakeholders?

References

- Abdulhamid, N. (2018). What is the relationship between alignment and washback? A mixed-methods study of the Libyan EFL context. [Doctoral dissertation, Carleton University]. Carleton University Institutional Repository. <https://repository.library.carleton.ca/concern/etds/mk61rh89v>
- Ahmed, A. (2018). Washback: Examining English language teaching and learning in Libyan secondary school education. [Doctoral dissertation, University of Huddersfield]. University of Huddersfield Repository. <http://eprints.hud.ac.uk/id/eprint/34799/>
- Alderson, J.C. & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13, 280–97. <https://doi.org/10.1177/026553229601300304>.
- Alderson, J.C. & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14, 115–29. <https://doi.org/10.1093/applin/14.2.115>
- Allen, D. (2016). Japanese cram schools and entrance exam washback. *The Asian Journal of Applied Linguistics*, 3(1), 54–67. <https://caes.hku.hk/ajal/index.php/ajal/article/view/338>
- Allen, D. (2017). Investigating Japanese undergraduates' English language proficiency with IELTS: Predicting factors and washback. IELTS Partnership Research Papers 2. IELTS Partners. <https://www.ielts.org/-/media/research-reports/ielts-research-partner-paper.ashx>
- Allen, D. (2023). English private tutoring and washback in Japan. In K.W.H. Yung & A. Hajar (eds), *International perspectives on English private tutoring* (pp. 37–54). Palgrave Macmillan. https://doi.org/10.1007/978-3-031-26817-5_3
- Allen, D., Yamamura, K., Meguro, S. & Nakamura, T. (2024). Washback from four-skills English exams in the Japanese secondary school context. *Language Assessment Quarterly*, 4–5, 404–426. <https://doi.org/10.1080/15434303.2024.2440895>
- Burrows, C. (2004). Teacher factors mediating washback. In L. Cheng & Y. Watanabe (eds) with A. Curtis, *Washback in language testing: Research contexts and methods* (pp. 112–128). Lawrence Erlbaum.
- Chapman, D.W. & Snyder, C.W. (2000). Can high stakes national testing improve instruction? Re-examining conventional wisdom. *International Journal of Educational Development*, 20, 457–474. [https://doi.org/10.1016/S0738-0593\(00\)00020-1](https://doi.org/10.1016/S0738-0593(00)00020-1)
- Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education*, 11(1), 38–54. <https://doi.org/10.1080/09500789708666717>
- Cheng, L. (2005). *Changing language teaching through language testing*. Cambridge University Press.
- Ferman, I. (2004). The washback of an EFL national oral matriculation test to teaching and learning. In L. Cheng & Y. Watanabe (eds) with A. Curtis, *Washback in language testing: Research contexts and methods* (pp. 190–210). Lawrence Erlbaum.
- Green, A. (2006). Watching for washback: Observing the influence of the International English Language Testing System academic writing test in the classroom. *Language Assessment Quarterly*, 3(4), 333–368. <https://doi.org/10.1080/15434300701333152>

- Green, A. (2007). IELTS washback in context: Preparation for academic writing in higher education. *Studies in Language Testing* 25. UCLES/Cambridge University Press.
- Qi, L. (2004). Has a high-stakes test produced the intended changes? In L. Cheng & Y. Watanabe (eds) with A. Curtis, *Washback in language testing: Research contexts and methods* (pp.171–190). Lawrence Erlbaum.
- Saville, N., & Séguis, B. (2025). Washback, impact, and consequences. In D. Allen (Ed.) *Washback research in language assessment: Fundamentals and contexts* (pp. 12–25). Routledge.
- Shih, C. M. (2009). How tests change teaching: A model for reference. *English Teaching: Practice and Critique*, 8(2), 188–206. <https://files.eric.ed.gov/fulltext/EJ859684.pdf>
- Shohamy, E. (1993). The power of tests: The impact of language tests on teaching and learning. National Foreign Language Centre, John Hopkins University. <https://files.eric.ed.gov/fulltext/ED362040.pdf>
- Shohamy, E., Donitsa-Schmidt, S. & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13(3), 298–317. <https://doi.org/10.1177/026553229601300305>
- Sultana, N. (2019). Exploring the alignment of the Secondary School Certificate English Examination with curriculum and classroom instruction: A washback study in Bangladesh. [Doctoral dissertation, Queen's University]. Queen's Graduate Theses and Dissertations. <https://qspace.library.queensu.ca/items/29801b52-8f27-4b89-94b3-0904aa35e673> .
- Tan, M. & Turner, C.E. (2015). The impact of communication and collaboration between test developers and teachers on a high-stakes ESL exam: Aligning external assessment and classroom practices. *Language Assessment Quarterly*, 12(1), 29–49. <https://doi.org/10.1080/15434303.2014.1003301>
- Tsagari, D. (2012). FCE exam preparation discourses: Insights from an ethnographic study. *Cambridge ESOL: Research Notes*, 47, 36–48. <https://www.cambridgeenglish.org/Images/22669-rv-research-notes-47.pdf>
- Tsang, C.L. & Isaacs, T. (2022). Hong Kong secondary students' perspectives on selecting test difficulty level and learner washback: Effects of a graded approach to assessment. *Language Testing*, 39(2), 212–238. <https://doi.org/10.1177/02655322211050600>
- Umashankar, S. (2017). Washback effects of speaking assessment of teaching English in Sri Lankan schools. [Doctoral dissertation, University of Bedfordshire]. University of Bedfordshire Repository. <https://core.ac.uk/download/pdf/151440224.pdf>
- Vyn, R. (2019). Promoting curricular innovation through language performance assessment: Leveraging AAPPL washback in a K–12 world languages program. [Doctoral dissertation, University of Iowa]. Iowa Research Online. <https://doi.org/10.17077/etd.iepe-7s2s>
- Wall, D. (2005). The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory. Cambridge University Press.
- Wall, D. & Alderson, J.C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*, 10(1), 41–69. <https://doi.org/10.1177/026553229301000103>

- 
- Wall, D. & Horák, T. (2011). The impact of changes in the TOEFL® exam on teaching in a sample of countries in Europe: Phase 3, The role of the coursebook; Phase 4, Describing change. Educational Testing Service. <https://doi.org/10.1002/j.2333-8504.2011.tb02277.x>
- Wang, J. (2010). A study of the role of the 'teacher factor' in washback. [Doctoral dissertation, McGill University]. eScholarship@McGill. <https://escholarship.mcgill.ca/concern/theses/xp68kg62v>
- Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing*, 13(3), 318–333. <https://doi.org/10.1177/026553229601300306>
- Watanabe, Y. (1997). The washback effects of the Japanese university entrance examinations of English-classroom-based research. [Unpublished doctoral dissertation]. Lancaster University.
- Watanabe, Y. (2004). Teacher factors mediating washback. In L. Cheng & Y. Watanabe (eds) with A. Curtis, *Washback in language testing: Research contexts and methods* (pp. 129–146). Lawrence Erlbaum.
- Yu, Y. (2010). The washback effects of school-based assessment on teaching and learning: A case study. [Doctoral dissertation, University of Hong Kong]. The HKU Scholars Hub. <https://hub.hku.hk/handle/10722/65273>

Appendix 5: Narrative summaries

Interview 1

Teacher 1 works at a SHS in a rural area in a prefecture in Eastern Japan. Her area has a diverse range of students, some of whom want to go to university but many who do not, or could not even if they wanted to because they are from poor families. Her school operates an integrated-course system so students can take more academic subjects for preparation for university or more vocational courses for entering the job market directly.

Regular classes involve pair and groups speaking activities, which occupy about 10–15% of the time. However, teachers sometimes struggle to get students to speak due to the class atmosphere and the fact that students' self-esteem is sometimes very low.

'Sometimes, it depends on the atmosphere that the classrooms have, students kind of hesitate to speak out sometimes, like, their self-esteem is really low, then they try not to speak out. So, we are struggling, encouraging students to speak a lot more sometimes.'
(1)

She works with an assistant language teacher (ALT) to design and implement speaking tests (one-minute, one-to-one interview format) and conducts 40 tests in a 50-minute class. The approach taken to teaching speaking, however, depends on the teacher.

There are six English teachers in the school and Teacher 1 has more progressive in her views towards English teaching. She feels fortunate that there are few 'younger' teachers who are willing to adopt the approach advocated by MEXT, that is, to teach English communicatively, but she also mentions two 'older' teachers, one of whom graduated from a prestigious high school and has contrasting views about English education. For instance, he believes that grammar should be taught before speaking.

'in [this] prefecture, my high school is a bit better than the other schools, I think, because out of six, three teachers are very young, I mean, they are 20s, so they are willing to learn, and they are making efforts, I think, trying to adapt many new approaches, so those teachers are hope, I think. But, on the other hand, one really old teacher who is from the more prestigious high school, his belief is that more, like, you have to learn grammar first before speaking out, so, yeah, I don't agree with him, but he has his own philosophy for teaching English, I think.' (1)

Also, the only time he changes his beliefs and attitudes towards teaching is when the exams change. For example, he changed his attitude towards teaching listening when she demonstrated that doing so raises students' scores on entrance exams (i.e., the Common Test). She believes he would similarly change his attitude towards speaking if the BCT-S was introduced as an entrance exam and consequently it would be easier to introduce speaking activities into classes.

'When listening activity introduced to the *Center Test*, I mean, the *Kyotsu Test*, I think I introduced more listening activity during the class, for the class. And then last, I did a

lot of listening activities, then students got higher score than the previous year. Then it's so obvious that their listening score became high because I introduced listening activities to the class, which had an effect on students and they got higher score, which is the good evidence, right? So, if speaking test introduced to the entrance exam, which means that we have to do that. So, it's going to be easier for me to conduct the class, including speaking activities.'

Other teachers would become highly motivated to teach speaking if it was part of the entrance exams. However, not everyone will be able to assess speaking. Teacher 1 told a story of how she asked the other older teacher to assess speaking as part of her mid-term assessment so that there was consistency across the classes. However, the other teacher's lack of ability and confidence in English speaking and evaluation of students' speaking led her to simply adopt the ALT's scores, rather than double-mark and discuss the scores with the ALT. That is, she 'gave up' trying to assign a score for each student's speaking performance. Therefore, teacher training is crucial if they are to teach, and in particular to assess, speaking.

She had previously taught in another city in the same prefecture for many years, and has only been in her current position for a few years. Her current school is more 'conservative' than her previous one. Although teachers in elementary and junior high school in the area appear more willing and able to implement communicative approaches to English teaching, senior high schools are more traditional. This is due to some of the senior teachers who are involved with writing the prefectural SHS entrance exam and have more traditional views about language education. The focus on accuracy, reading and grammar she noted is greater in her local exam than in those of other prefectures, and this, she believes, is one reason why students are afraid to speak up in class. Similarly, she noted that the third-year students who were applying to universities were taught with a focus on entrance exams, whereas those who were planning on going to technical colleges were taught differently: 'we will have a lot more fun type of activity, like, watching YouTube, and then, like, listening to songs and that kind of activity.'

Although she had some questions about how it was scored and why some elements were included, she felt it was an appropriate test in terms of the relationship with the Course of Studies and her own teaching. The tasks overlap with her teaching, especially Tasks 1 and 2. However, few third-year students would be able to handle Tasks 3 and 4 due to their limited proficiency.

If the test was introduced, she would spend more time working on developing students' spoken production ability. In particular, she would spend time on activities that involve planning and organizing speech, and creating coherent arguments, and also focusing on 'higher level' cognition and 'critical thinking'. Currently, she mainly instructs students on how to produce well-organized writing. She would also do this for speaking to help students cope with the task demands. Moreover, she would instruct students in comparing, analyzing and evaluating. Currently, her speaking activities and assessments are more focused on situations, such as going to a restaurant, which she characterizes as 'shallower' in terms of cognitive demands.

Thus, the BCT-S would stimulate change in the kinds of speaking activities and assessments that she employs.

She is concerned about the time that she has available to cover the official textbooks and the material therein and suggests therefore that the textbooks would need to change, too. She would need to teach speaking more but also would need more time for teaching vocabulary and for having students prepare scripts for speeches, which would increase the amount of time required for writing. Although she recognizes that students need more training in speaking spontaneously, some teachers think speaking requires planning a script first, so they may do more planning-type activities instead. She also thinks some teachers would not cope well themselves with the BCT-S due to their low proficiency in speaking and believes they may instead come up with an excuse such as ‘you need to know grammar when you speak properly’. In other words, there is likely to be some resistance to teaching speaking due to teachers’ own abilities and their beliefs about teaching English. Moreover, without appropriate training in assessing speaking, some teachers will only focus on accuracy, which could make students stressed when speaking in class and during assessments.

Interview 2

Teacher 2 works at a public senior high school in an urban area in Eastern Japan. The school is academically mid-level but provides specialized English education. Students study seven to eight classes per week across the three years, which include the official classes recommended by MEXT and additional classes focused on reading and presentation. The teacher has extensive teaching experience teaching English in high schools and cram schools. He is involved in textbook writing and is a speaking examiner for the EIKEN suite of tests (e.g., Grade Pre-2 and Grade 2) produced by the Eiken Foundation of Japan (henceforth, these tests are referred to as EIKEN). He previously worked at a prestigious high school and sometimes refers to this as a source of comparison.

The students are average or slightly below average in terms of English ability but they are ‘active towards using English’. Classroom activities involve input and output such as readings followed by pair work speaking activities. Activities are generally similar across the three years but the topics develop from more concrete, personal ones to more social, abstract ones. In the additional classes, students present once a month to the class. In other classes, pair work speaking activities are assessed approximately every other week. One-to-one speaking assessments are also conducted at the end of each of the three regular terms but this depends on the teacher who is in charge of that year’s cohort. Overall, the teachers in the school are very willing to teach English speaking in class and so there are a lot of oral activities.

If the BCT-S was introduced he would significantly change his way of teaching so that his students could get good marks on the test. This is because his current teaching is ‘not always

compatible' with the test. He would reflect carefully on what he would have to do to help them get 'a good score' on the test.

He emphasizes that currently a key goal for teachers in his school is just to get students to talk, that is, developing some degree of spoken fluency. He characterizes his students as shy: it's 'hard work for the English teachers to have them speak'. He sees this focus on fluency as a key characteristic of teaching at his school and that this leaves less room for teaching other things, such as logical structure, linguistic accuracy, grammar. He would 'really have them care about what they say', rather than just saying something. His students now often answer questions with words but not sentences, so he would get them to focus more on producing logically-ordered sentences. Thus, the 'instructions' that he gives to his students would be very different.

He thinks that it is not clear what students need to focus on to get a good score on the BCT-S.

'the explanation about assessment is very general it's like oh this vocabulary is not rich so this is B1 and some mistakes, but well that's good content so B2. It's very general explanation of assessment. So, it's tough for the Japanese students to know which, you know, how do I say, what to focus on? What do I do? At the same time, the teachers will be confused about what we focus on, teaching vocabulary or structure or logic, you know, it's I think it's not for me, but for most of the teachers, I believe, it's very difficult for them to get this, you know, how to teach in order for the students to get that good score on the test.'

Japanese students and teachers like to know exactly what to focus on, and this is one reason why EIKEN is popular. For example, teachers know that test takers must mention specific things from the prompts, take care about their use of tense, and use pronouns after mentioning the character's name for the first time. These specific 'instructions' can be given to students. In contrast, teachers will struggle with knowing how and what to teach if the BCT-S is introduced.

He would do more picture description task but he feels it would be 'very hard' for his current students. He says they would organize more speaking activities for students and in general would do similar activities to those in the test. Teachers would be motivated to teach for the BCT-S and many would want to teach speaking for the test, which he sees as more 'authentic' than EIKEN, where students are just following rules and doing what they had prepared. However, it is important to notice the conflict here: teachers would like to teach for the more authentic (i.e. less formulaic) BCT-S, but they also want to know exactly how to guide their students to get a good score.

Another conflict appears in his assertion that teachers have two goals: using English in 'a globalized world' and passing exams. Teacher 2 characterizes the former as 'practical skills', which include reacting to questions and 'extending their thoughts or experience', and the latter as 'test-taking techniques'. It would be easier if these two goals became one (i.e., develop English speaking proficiency) but he sees the Japanese culture of English teaching as different from that of the 'universal English teaching world'.

He says students will be motivated to study speaking if it enables them to get better scores on the BCT-S. To support this claim, he says currently most of his students' motivation for speaking in class derives from the desire to get good scores on the EIKEN test, which they can use for university entrance. Therefore, the BCT-S would replace the EIKEN as the source of motivation. However, while EIKEN is coachable, the BCT-S is not, at least not yet. He says 'we can find a coachable way to teach but most of the teachers do not know how to prepare for this kind of test.'

He noted some of the questions were unsuitable for his students, that is, most students would not know how to answer them properly and would give overly simple responses (e.g., 'What is your favorite month of the year?' would elicit responses such as, 'February, my birthday') or they would all answer in the same way (i.e., 'describe a challenge' would elicit the same response of 'entrance exams'). This he sees not simply as a fault of the test but of how his students do not have the ability to extend their responses to a variety of question topics. However, he says that students can respond to other questions (e.g., 'tell me about a difficult experience you had in your club activity') because they are familiar and used in interview tests for junior high school entrance.

He says that the BCT-S is compatible with the Course of Study, but that in reality what they do in the classroom is 'not always consistent with the Course of Study'. In fact, he says many teachers have not read or do not really understand the Course of Study, and instead they just teach in 'their way'. He said that domestic textbooks are not sufficient, which is why he currently supplements his teaching with foreign-published, 'non-official' resources. Such resources would likewise be necessary if the BCT-S were introduced, and if students got used to these materials, they would like them. He thinks his teachers would not need any specific training but perhaps some would have to 'train themselves a little more'.

Regarding the semi-direct format of the test, he says that compared to a decade ago students are very used to taking tests on tablets and that some even prefer it to a direct-format speaking test. No special preparations would be necessary and 'no different approaches' would be needed. Regarding cram schools, he does not think that they have nor would change their approach to teaching: They've been teaching the same way for decades.

Interview 3

Teacher 3 works at a senior high school in a suburban area in Eastern Japan. The school is a vocational school that specializes in programming, information processing, and bookkeeping. Fewer than half of the students go to university and most of those do not sit the Common Test or the competitive general route exams at universities; instead they take alternative entrance routes involving essays and interviews. Students take the minimum number of courses in English (i.e., English Communication 1–3 through the three years, and Logic and Expression is elective. Only nine out of approximately 350 students across all three years currently take this elective course). The teacher says they are 'very good students, but their interests are not

English.’ In other words, many students are not motivated to study English and so the teacher pays attention to keeping them interested through gamification and use of educational technology, which appeals to the students.

English lessons typically begin with a ‘small talk’ speaking activity followed by a focus on the textbook, which is one of the official textbooks recommended by MEXT. The book is largely focused on reading and includes few speaking activities. Speaking occupies between 15–20% of class time. There is a performance test every term that all three English teachers conduct: the test is a one-to-one interview format in which students describe a picture and answer questions. Students are generally shy and unwilling to speak in front of class, but they are able to speak with their peers.

Students could cope with the first two tasks of the BCT-S but not the second two tasks. If the BCT-S was important for students, the teacher would change teaching content rather than methods, and increase the amount of time focused on speaking activities because currently there is too much focus on input and comprehension. She feels students are familiar with the kind of Q&A task in Task 1. She would conduct picture-describing activities (as in Task 2) more frequently, and provide instructions and sample answers for students. This task is already done in the performance test and is therefore aligned with current assessment practices in the school. Task 3 (comparing pictures) is unfamiliar to students as she has not done that kind of task in class. She has taught the grammar for comparing but feels students cannot use it and so this task would be ‘very difficult for many students’ including third-year students. She says her students have done a ‘long turn’ task where students prepare for a short period then speak to the teacher (i.e., as in Task 4). They did this task in an assessment but it did not work well. Students could not speak enough and ‘they lost confidence to speak’ so the teacher changed the format and made the regular assessment easier. Students could not create arguments to ‘persuade’ and this was because ‘maybe my teaching style was very bad’ but also that ‘maybe the lack of practice affected [the outcome]’. However, students also do a presentation test where they have time to prepare out of class and then present for 1–2 minutes using slides or movies.

She could set aside time for more speaking activities if she could re-organize the curriculum. Classroom activities currently involve more interaction (i.e., pair work) than production (i.e., monologues) because production tasks involve a lot of preparation time, which is difficult to secure in the curriculum. Therefore, she would do more interaction activities but she would also have to intentionally set time for production activities.

‘Interaction will be the good practice to do the speaking test. So, they will be motivated, students will be motivated through doing this speaking test. And they will notice the importance of interaction activity in daily lessons.’

If she does not prepare students for the BCT-S, they may lose their confidence or ‘give up studying English’. However, if she teaches ‘tips for doing this test, maybe they will engage in the activity, they will engage in the test. They will maybe get a better score’. To do this, she would

introduce basic questions in the first grade and give them time to prepare their own and read other good examples. Then, she would increase the difficulty of the topic in the second grade, and finally use the same kind of tasks as in the test in the third grade. She would focus on the topic, increasing the difficulty by using less-familiar topics at higher grades, but she would also reduce the amount of time given for preparation as students progress.

She feels the BCT-S is aligned with the Course of Studies, which recommends speaking performance tests. It would be aligned with teaching ‘if the topic is relevant to my students and also topic is related to that textbook content’. However, she also says that it is most relevant to first- and second-graders, whose textbooks have presentation activities and expressing ideas, while third-graders focus more on reading skills. She says that she may need to do a special lesson aside from the textbook-focused classes or that she could possibly integrate speaking into the regular classes. This planning would take time.

She would need to acquire additional materials as her single textbook is insufficient. She can currently access resources online (i.e., good examples of speaking on YouTube or in movies) to help with preparation. In addition, she would like access to online resources that give examples of relevant speaking activities in the classroom. Also, she thinks she would need to find a way to evaluate the students’ performance using more BCT-S-like questions but keep the MEXT-consistent assessment marking scheme (i.e., A–C, expression, knowledge, and attitude).

The format is suitable for her students who are shy to speak in front of others but who like using computers. They have recorded their responses and submitted them to the teacher in the past, so the BCT-S format would not seem too unfamiliar to them.

Although most students would become motivated if the test was important, only a few students who aspire to work in English may be motivated by anything other than achieving a sufficient score on the test. The other two English teachers would be able and willing to teach speaking more if the test was introduced but ‘the elder teacher has difficulty with using internet’ so maybe she would need some assistance. Also, she would need to discuss and gain consensus with these teachers about changing the curriculum. Getting the others on board and revising the curriculum would be a challenge, but it would certainly be possible.

Interview 4

Teacher 4 works at a private co-educational six-year secondary school in a largely-rural prefecture in western Japan. The school is high-ranking in the prefecture and students typically come from wealthy households. Most students aim to gain entrance to a top-ranked university and are prepared to study hard, and parents have high expectations for their children’s academic achievement.

‘My school is a private institution, and students’ and parents’ expectations are really, really intensively focused on getting into good universities, so-called good universities. So, their expectations to get good quality education, and English, one of the major subjects, and because of their expectations, our lessons focuses on preparing them to

pass the highly prestigious universities. They are quite ambitious, so they are aiming at Tokyo University, Kyoto, Osaka, Nagoya, all those high-standard universities.'

The school is large with over 1,600 students, 900 of whom comprise the student body of Grades 10 to 12. Thus, regular classes have 50 students and only special courses, such as an intensive writing class taught by Teacher 4, have smaller class sizes of around 20 students. There are 12 Japanese-L1 English teachers and three English-L1 teachers. Students each have access to a personal tablet-computer device though these do not currently form a major part of their classroom-based learning of English due to the difficulties of using technology with large class sizes.

In line with students' and parents' expectations, teachers are highly focused on preparing them for entrance exams and this begins in Grade 7 and intensifies in later grades, especially Grade 12. Given that the goal of passing the high-level university entrance exams is the priority for all stakeholders, the contents and methods of English language teaching are geared towards success on these exams and the decisions about what and how to teach are guided by teachers' understanding of the requirements of such exams.

In senior high school there are six classes per week. In Grades 10 and 11, three classes are focused on reading, two on grammar, and one on listening and speaking. In the latter class, there is a speaking test which takes the format of an interview, a debate or a speech, but which is 'a kind of memory test' as 'there is no spontaneous conversation'. In Grade 12, the classes are similar but there is no listening and speaking class, and teaching is oriented towards students' university choices (e.g., science vs. humanities courses). There is also a listening class that focuses on preparing students for the listening part of the Common Test.

The style of teaching is described as 'old-fashioned' and 'traditional'. It is 'lecture-based and there's not much interaction involved'. The teacher lectures while students take notes and revise them for the mid- and end-term exams. The focus is on 'improving grammar knowledge and reading comprehension, and translation from Japanese to English'. This style is similar across grades and teachers albeit with some variation. For instance, in some classes, Japanese-L1 and English-L1 teachers team-teach more 'conversational style English' and one British teacher tries to incorporate 'freestyle conversation' but 'the time is really limited' to around five minutes in a class. Teacher 4 tries to incorporate more pair-work activities, though these typically have a grammar or vocabulary focus. There is 'not much spontaneous conversation' and pair work is not a common feature of classes.

The lecture-based method is partially justified based on the large class sizes, for which it is 'the best possible way'. However, the primary justification is the pressure of students' and parents' expectations:

'We always feel some pressure from students and parents, because as I told you in the beginning, their expectation is to getting into a university. And we want to meet the expectations, of course, obviously. And we need to adjust our teaching to that.'

Teacher 4 is a confident user of English who feels a conflict between what she would like to do and what she needs to do. She has experience of supporting students on overseas study programs and says that:

I know they want to improve their speaking abilities, but in a limited amount of time, length of time and environment, in regular class hours, and weekdays, it's not possible to make it happen. So, for me, as a teacher, it's sad, but trying to improve their speaking abilities is a kind of extracurricular activity.'

Ultimately, alongside her appreciation of the benefits of learning English speaking, she feels the entrance exam system is the key reason why English is taught the way it is at her school:

I've been a learner of English my whole life. So, I know how wonderful it will feel if my students would be able to speak English fluently. And I've seen lots of them have reached that level. And it's wonderful. And I would like to have lots of students enjoy that. But it's not possible under this current educational system, because the university entrance exam is the same.'

If the BCT-S was introduced, Teacher 4 believes that 'it's going to be a massive change.' She would use previous tests and sample materials and have students 'practice as often as possible'. She believes she would reduce the time spent on teaching grammar, composition and translation, while increasing the number of speaking activities in class. She feels that she would need to start from the basics 'like pronouncing each word correctly' because 'they basically don't have much experience to express themselves in public. So, it's really high pressure for them to speak out, speak up in class.'

She believes class sizes would perhaps need to be reduced in order to promote a supportive atmosphere for speaking in public. However, for this the number of teachers would need to increase and these teachers 'who are ideally fluent in speaking English. Maybe more ... native speakers of English.' Meanwhile, the school may be too small to accommodate so many smaller-size classes and thus 'maybe we would teach in a corridor or maybe in a garden or somewhere'. When prompted to consider the situation where class sizes remained the same as they are now, she replied:

'we would have to find different kinds of tasks rather than just teaching grammar and vocabulary and translation, right? ... Maybe we would increase the amount of speaking activities, speaking tasks within the class period and we would have to give up lecture style sometimes and we would increase more pair work, group activities and yeah, we would have to change lesson plans from traditional lesson style, lecture style lesson plans towards more communicative style.'

She suggests there would be concern from teachers in other subject-departments who worry about a concurrent drop in reading or writing ability if communicative lessons were adopted. These concerned teachers, she believes, have a 'very old mindset that English skills can be only learned from reading and writing' because 'that's the way they were taught as students'. Therefore, 'we would have to make sure that the speaking ability is part of their English four

skills and we should make sure that we wouldn't lower the students overall English level just by increasing the amount of communicative lessons.'

She sees the overlap and potential for teaching speaking and writing together, giving Task 4 as an example. She could use this task in her writing lesson but students would need special training because they haven't had opportunities to speak. She says that she could incorporate speaking tasks into her writing classes if the topics overlap (e.g., online vs. face-to-face communication in Task 3). She believes new materials and online resources would be needed because current materials are insufficient to teach speaking. She thinks individually-tailored learning, such as the use of individual tablets, would also increase.

Currently, her students are not motivated to speak English. However, their motivation to study speaking would definitely increase if they had to take a speaking test: 'They would do anything to pass the entrance exams'. She refers to a number of her students who had to take the EIKEN Grade Pre-1 test, for which she helped them prepare for the speaking test, as an example of how they will study speaking if they are tested on it. At the same time, students are under a lot of pressure, and she feels a speaking test would increase that pressure and make them stressed. Because they have family financial support, they may go to conversation schools or take online lessons with native speakers to help them prepare. This additional burden would possibly create more stress. On the other hand, she notes that if the directions for the test are presented visually as well as aurally, that would reduce the pressure because learners can read and understand what is required. Also, if the test scores were made available to students, that 'would probably become part of the motivation to get tested and the results would be motivating for some students'.

In terms of test difficulty, she believes Task 1 should be quite easy for her students, while the other tasks would be more challenging. Students have difficulty creating longer, cohesive narratives, which would make Tasks 2 to 4 more difficult. They would need special training. Nevertheless, her students would be able to cope with Task 4 if they had practice.

Although Teacher 4 could teach speaking to the level required, only around half of the teachers would be able to do so if the BCT-S was introduced now. The other half would struggle because they 'don't enjoy speaking English so much'. Not all of the teachers have had overseas experience and they currently do not speak English in class; therefore, they would have difficulty producing English to the level that they would want their students to be achieving.

Teacher 4 has a number of concerns related to the introduction of speaking tests. Firstly, she is wary of ministry-led changes in English because of the failed reform proposal to introduce four-skill exams. Moreover, she believes English teachers are singled-out by ministry-led initiatives. They always have to change to something new, and this often conflicts with students' or parents' expectations. She is also somewhat cautious of speaking tests because sometimes they are not fair. Although she believes that if it were logistically possible it would be good to have a speaking test for university entrance, she acknowledges that this is

contradictory to the requirements of top universities: ‘academic universities ... want to develop scholars who can read and write good thesis and who can get involved with research. So first and foremost, what they want is, I would say, reading and writing skills.’ She is also not convinced by the necessity for students to learn speaking in high school:

‘I think speaking ability would come later. After they have mastered reading and writing, speaking skills would come later in life. They can start after they can start developing their speaking, making presentations after they go on to graduate school ... they don’t have to start in high school.’

This belief may be related to her own experience in which she had started learning English conversation seriously in her final year of university. She learned the phonetic alphabet and after that her English improved dramatically.

Interview 5

Teacher 5 works in a senior high school in a largely rural prefecture in central Japan. The school has six full-time teachers of English, one part-time teacher, and an ALT. Two teachers from the attached part-time school also assist in the main school. There are roughly 350 students who come from the surrounding area and who vary significantly in their abilities and future goals. From Grade 11, they are divided into either normal or advanced ability-level courses according to whether they plan to go to university, and science or humanities depending on their preferred subjects. Students from each course are placed in one of three levels for their English classes by using a mock exam created by Benesse. Teachers prepare different classes and use different textbooks according to the class level. Because teachers rotate their teaching duties each year, they need to become accustomed to using various materials for the different grades and ability-levels, which is a challenge for teachers.

English classes are divided into the official courses of English Communication and Logic and Expression. Teacher 5 teaches Grades 11 and 12 and has not yet taught Grade 10. Describing English Communication, she says:

‘we use an English textbook and we read the textbook and the teacher will explain some grammar points and then try to do some activities based on the topic. But for third years (Grade 12), we don’t use textbooks so often. We use some textbooks for entrance exams and we basically practice reading very long sentences and solve some questions.’

English speaking activities occupy about 10% of class time during Grades 10 and 11 but this goes down to 0–5% in Grade 12, when entrance exams dominate. Speaking assessments, such as one-to-one interviews, prepared presentations, and show-and-tell, are often conducted with the assistance of the ALT in English Communication classes during Grades 10 and 11. In general, speaking-focused activities and assessments are rare for Grade 12. Instead, writing

activities are more common in the Grade 12 Communication classes and the ALT also assists with these.

In the Grade 11 Logic and Expression classes, which are referred to as the 'grammar class', students review all of the grammar they learned in the first year. In Grade 12, grammar textbooks are used and students practice translating from Japanese to English (but not English to Japanese):

'students solve some problems in the workbook at home. And in the classes, teachers will explain the answer in a very old-fashioned way of teaching... [students] don't speak so much in logic and expression classes.'

The curriculum for each grade is typically organised around reading, grammar, and composition, while speaking does not feature prominently. There appear to be a number of factors that influence what is actually taught in the school. Most pertinent are entrance exams, which are emphasized as guiding choices in instruction in all grades, especially in Grade 12. For instance, there is a specific listening class in the Grades 11 and 12 that focuses on preparation for the listening part of the Common Test, and English Communication is primarily geared towards exam preparation in Grade 12. For Teacher 5, when describing an advanced level class, it appears that her confidence in her teaching ability, students' goals and student's interest all influence her decision not to teach speaking so much in class:

'I sometimes hesitate to provide speaking task because I feel I'm not so confident about managing speaking activity. And some students look very bored or they, because it's an advanced class, they want to pass the entrance exam. So, they feel like speaking activity is not so important for passing entrance exam. So, I hate that kind of atmosphere. So, I'm, I feel hesitant to do speaking activity [...] but for second year extra communication class, every time I do warming up activity, speaking activity. But some students seem very, seems to enjoy speaking activity, but others' facial expression is very like plain and yeah, I can tell they don't enjoy.'

Regarding other teachers, she does not know what they do in their classes and what is taught depends on the teacher, especially regarding teaching speaking, for which there is no specified curriculum. She feels that generally the teachers believe in the importance of speaking and 'we try to make classes communicative to some extent'. However, she mentions an extreme case of a former teacher who completely rejected the idea of teaching communication and speaking in class: 'He said he hated communication'. This teacher did not believe speaking was important and instead focused on accuracy in language use, grammar and composition. He was very focused on exam preparation: 'he gave some entrance exam questions to students a lot, even in second year'.

'So, for some, some experienced teacher, they, I think they want to focus on more grammar or reading. Hmm. So, he moved to different school this year, but one of the teachers in last year, he, like, hated communication class and he didn't do any speaking activity or pair work. He didn't believe that speaking is very important ... he

said he hated communication ... he put importance on accuracy. ... so, he wanted students to write very accurate English sentences. He wanted them to speak very accurate English. Hmm. And he believed that basic grammar knowledge is very important. So, he liked to teach grammar or English writing rather than speaking.’ (5)

Each year, two teachers supervise one grade, and the more experienced teacher leads in deciding the curriculum for that year. Teacher 5 feels a little frustrated with the way the curriculum is determined and often talks to another young teacher about the situation and would like to change it, but she is not sure of other teachers’ opinions. She says it is hard to tell her opinion to other, more-experienced teachers. In a rather exceptional case, she has a ‘free class’ in which she can decide the content and plans to try English debates or presentations.

‘Young teachers, teachers in 20s, there are two, me and another teacher is in 20s and we want to change this situation. But, I don’t know other teachers’ opinions or, we feel very hard to change this situation. So, there are two leaders for each year, two teachers for first year, two teachers for second year. And always, I, my partner is very experienced teacher and it’s very difficult to tell my opinion to the other teacher. So, yeah. But, yeah, many times I and the other young teacher, I sometimes talk with the other young teacher and about how we can change this situation.’ (5)

If the BCT-S were introduced, she believes that speaking would be part of the planned curriculum and all teachers would increase the number of speaking activities. This curriculum would be based on the content of the exam, starting with the first task. She describes the situation for teaching listening in Grades 11 and 12 and imagines that teaching speaking would follow a similar pattern: ‘We practice listening by using listening workbook, exercise book, once a week. So, similar to the listening, maybe we would introduce speaking and practice once a week or half class every week.’

She would continue her warm-up speaking activities, which sometimes include Q&A or a picture description task and which are always interactive. She says that these interaction activities are important for developing the ability to ask follow-up questions, which connects to the ability to keep talking (i.e., adding more information).

Her advanced students could cope with the first two BCT-S tasks, which are familiar to them, but lower-level students would struggle to come up with reasons in their responses. Although they have learned linking phrases and relevant grammar, students have never done a comparing-pictures activity (Task 3) or a long monologue (e.g., Task 4). Her students sometimes do short speaking activities related to social topics covered in the textbook, which overlaps with the test tasks, but she would expose them to a ‘wider variety of social topics ... because in speaking test, background knowledge is very important. So, if they know a lot of things, they can talk about them’. Her current materials would need to be supplemented because they only cover ‘10 topics in a year’. Also, she believes students would improve quickly. She cites her students who come to her to practice for the EIKEN test and who, once they have been given some tips and start practising every day, improve quickly.

Her students would be motivated if there was a speaking test: 'If the speaking activity follows the test, they would be motivated.' Also, teachers would be motivated. She describes how the teachers are influenced by exams through an example of how they are preparing for the Common Test: 'I feel our way of teaching is affected by entrance exam a lot'.

Although she could give advice about major and general areas for improvement following speaking activities, she is concerned about being able to give individualized feedback, like she does for writing. She feels that recording may be difficult to do each time, but students do have their own tablet so recording is a possibility and they are not unfamiliar with tablet-based assessment because they have taken the GTEC test before (all Grade 11 from this year) (GTEC is a suite of English four-skills tests produced by Benesse Corporation). She is not confident that she could assess spontaneous spoken production in a performance test using multiple criteria like those used in the BCT-S. Although she has used multiple criteria to assess presentations, the speeches were prepared and so accuracy was generally quite high. She is unsure about how to assess a balance of accuracy and fluency during spontaneous speech.

Interview 6

Teacher 6 works at a vocational school in a rural prefecture in Western Japan. The school has roughly 900 students, the majority of whom do not go on to university, while those who do tend to apply to local universities. There are six English teachers and a full-time ALT. Teacher 6 previously taught at an academic high school in the same region and moved to this new position a few months ago. Although she is still getting used to the system there, she notes the difference in terms of academic ability and future goals. Importantly, most of the students seek employment after high school and so in Grade 12 they are focused on securing employment, rather than entrance exams. Therefore, while school grades may be important for them, they do not need to take any university entrance exams, or English exams. Therefore, students are generally not motivated to put effort into learning English.

The English level of the students is low compared to those at academic high schools. Students take three 'English Communication' classes per week. Logic and Expression, the other MEXT-designated course, is only taken in Grade 12 as an elective course for those students wishing to improve their English, typically for university entrance exams. This year only around 12 students are taking the course.

Although they are low level, have a limited vocabulary, and have difficulty constructing basic sentence structures in speech, many of the students like communication. They enjoy interacting with the full-time ALT, who teaches speaking classes alongside a regular teacher every other week. Their responses are typically very short, and often one-word responses: 'students are pretty good at communicating ... one-word answer, or, you know, reacting'. However, they do not spend enough time studying English in school and have little motivation to improve beyond a very basic level. A handful of students, on the other hand, come to Teacher 6 for speaking practice for the EIKEN test, and they are very motivated to learn.

Although there are 40 students in non-English classes, these are divided numerically into two groups of 20 for English classes. English classes typically begin with vocabulary activities, followed by a short reading and comprehension questions and a personalization question from the textbook, further short exercises and then some longer output activity using a worksheet adapted from the textbook content. Pair work is common in class. Grade 10 students are happy to speak more, having done a lot of it in junior high school. The textbook they use in the school is 'little bit easier than their junior high school... so they talk a lot... so they like speaking, and they seem to enjoy pair work.' Spelling, on the other hand, is difficult for them. Grade 12 students, on the other hand, do not seem to like speaking in class.

All the teachers use the same textbooks, handouts, and tests. The content of the worksheets, and the mid-term/end-of-term tests, is determined by the teacher in charge of that year. A vocabulary textbook is used for learning and regular assessments, which Teacher 6 says resembles the method often seen in other, more-academic schools. Speaking assessments are conducted in Grades 10 and 11 (e.g., show and tell), but not in Grade 12.

Teacher 6 found it hard to imagine a situation where her students would need to take the BCT-S as a high-stakes exam. This is because the students typically do not take any English exams or university entrance exams and therefore 'teachers are not teaching English for the test'. Consequently, introducing a high-stakes exam is in itself a break from the norm and would require considerable rethinking of the way English is taught in the school.

Crucially, teachers, students, and other stakeholders would need convincing of the importance of English speaking. Because it is a vocational school, non-English teachers would need to change their view that English is important, in addition to finding a job, which is currently students' main goal. These teachers would turn to the English teachers to prepare the students. For students, 'probably the first year, we kind of like religiously say speaking is important.' Also, to motivate them it would be necessary to include 'some kind of harassing part of class. If they can't do this, then they have to stay after school. Otherwise, lots of them just ignore speaking.' For parents, 'maybe we would tell parents how important it is and make sure they all go for speaking, you know, because that kind of mentality doesn't exist right now.'

Teachers would need to create a speaking curriculum, which would focus on different tasks in the test. She imagines teachers may 'make them prepare for this type of question. Just make them memorize this model answer or something like that. Then ask them to write on the written test and stuff like that, maybe.' Teachers would also need to analyze mock test results to plan classes.

A major issue is the fact that students only have three classes per week and they currently need to cover the official textbook content, which is low-level and focused not only on speaking. Therefore, with such a limited time the teacher feels it would be very difficult to teach students speaking for the BCT-S in addition to covering the textbook content. Vocabulary building would need to be reconsidered as the current supplementary vocabulary book is based on traditional entrance exams. The new vocabulary book would be used to plan

classes. 'And we may reduce the stress on spelling. We may change our vocabulary test based on just the speaker or just a quickly response or something.'

Classroom management would also be an important issue. Pair work would need to be considered carefully 'because lots of students don't take it seriously... Like the strong boys and weak boys. Strong boys don't want to talk. Then pair work doesn't work ... lots of punk type of students don't want to do it'. This issue is especially important in Grades 11 and 12, while Grade 10 students are better at working in pairs and communicating in class, at least at the beginning of the year (in 'the summer [Grade 10 students] are becoming like, you know, acting out and kind of getting rougher'). They may need to ask for more teachers due to the difference in levels and attitude and the need to separate students in order to make pairs work: 'Some students who can do it or who want to do it have to be separate.'

Currently, students would be able to attempt the Task 1 but would struggle with the other tasks. One reason is that they are not used to extending their ideas in English. They are used to restricted output activities, following set patterns provided by the teacher in class. She says, 'they should be able to create a sentence by themselves. But we kind of don't do that. But we may have to do that. Seriously. Teaching speaking is a little bit, well, lots of teachers are just showing the pattern. And they really want the pattern. And that's good. That's good to learn the pattern. But it's not only that. So, the good students may have to move on to the new level.' She continues, 'some students who are keen, who want to learn and who want to improve or whatever the reason, they're exposed to, you know, native speakers English or something, then they can go ahead, go to the different mode. But most students don't have that at my school.'

Task 2 (picture description) should be achievable if students have the vocabulary and sufficient practice because it is similar to EIKEN practice. This may be covered in the textbook but she is not sure because she is still not familiar with the materials. Regarding Task 3, 'comparison can be interesting enough. And if they get the how to answer and vocabulary learning with vocabulary, then they can deal with it. ... But then they when they have to express their opinion, then it is really difficult.' She describes the ability level of students, which includes thinking in Japanese and being unable to express their ideas in English. Regarding Task 4, 'they have to decide on what they have to say, right? The time is so short, and I don't know if they keep taking the test like that'. She believes that her students are not good at taking tests and that it would be difficult for them to prepare for and succeed on the test: 'good test takers, I mean, skilful test takers have gone to other schools.' However, the format would not be a problem as students each have a tablet, and although recording speech is not a feature of regular classes, some may have experienced tablet-based English tests.

Most teachers are 'good at teaching' and would be able to adapt to teaching speaking for the BCT-S. However, 'but some teachers are not really good at speaking, maybe. Yeah. So that's, that's a difficult part'. Not all teachers have a qualification showing the required level of speaking ability (EIKEN Grade Pre-1) and until recently there was a teacher who only taught translation. Some teachers may also feel it unusual not to have a read-aloud task in the test,

which features in the well-established speaking tests of EIKEN and GTEC, and is also used in some schools for assessment purposes.

Interview 7

Teacher 7 works at a co-educational public senior high school in a suburban area of a prefecture in western Japan. The school is middle-ranked in terms of academic achievement and the majority of students go to university. There are roughly 700 students in the school and each grade has six classes of approximately 40 students. One of these classes is a specialized English course in which classes are conducted completely in English and students take part in an English Camp and other language-related activities. This course has 40 students, who are selected through a separate entrance exam and who are higher level at English than students in other classes. Teacher 7 is in charge of this specialized English course, as well as Grade 10 students, and he conducts most of his classes completely in English.

There are ten full-time teachers, two part-time teachers, and two full-time ALTs. The official curriculum includes three English Communication and three Logic and Expression classes. These classes follow the official textbooks closely and alternative textbooks are not used. The content and pacing of lessons is similar across teachers, though differences in teachers' confidence in their ability to use English and also technology create some variation across classes. Although the difficulty level increases gradually, the content and style of teaching does not change significantly across grades. The specialized course includes additional classes focusing on debating and conversation.

A typical Communication class follows the textbook and begins with a topic lead-in, followed by a listening activity, then a reading activity and comprehension questions. Students then typically do a story-retelling activity that they record using their tablets and submit to the teacher for archiving. Logic and Expression classes also follow the textbook and involve the four skills and a lot of pair work. Each class has a main theme introduced through reading and listening tasks, and a focus on a particular grammar point, which the teacher clarifies in class. This is followed by speaking and writing activities. Speaking performance tests are conducted every semester, such as show and tell and self-introductions in Grade 10, and a poster presentation in Grade 12, which is recorded and evaluated.

Overall, Teacher 7 believes that if the BCT-S were introduced, his teaching would not change dramatically:

I actually teach students speaking a lot probably compared to other high schools. My high school's standard deviation is 50. But in my opinion, my students' speaking ability is quite high, like a standard deviation of 60 or so, because we put more focus on speaking in the class. So usually in high school, they don't have much chance to speak in the class. But actually, we already gave them a lot of chance, the opportunities to

... speak. So, I don't think that I have to change the way of teaching at my school so much.'

He says that he would not increase the amount of speaking in class because students already study speaking enough and increasing the amount would negatively affect the balance across the four skills. Also, he says the materials that are currently used, the official textbooks, are sufficient to teach speaking. He refers to the Logic and Expression textbook which includes speaking activities on social topics, as an example of how the books cover the task requirements. On this point he notes that in other schools teachers may skip over speaking sections and focus on grammar textbooks instead; however, at his school, teachers do not skip those parts and thus students have plenty of opportunities to develop their speaking ability.

In relation to this, he is also positive about the idea of introducing a speaking test for university entrance, precisely because his students spend significant time and effort studying and practicing speaking in English: 'Balance is important. Actually, I spent too much time for speaking in the current situation. So, I want to, the university exam to be changed. Yeah. My students will get that advantage.'

Regarding motivation, he believes his students are already intrinsically motivated to speak English and they enjoy it: 'Students enjoy speaking. So, they are already motivated inside.' If the BCT-S were introduced, they would also become extrinsically motivated: 'So from both sides, they will be motivated. That's good.' On the other hand, he says that some students don't like speaking English, but they would be motivated to do so if speaking was part of an important exam: 'those students are especially motivated by the exam'. Meanwhile, he thinks a small proportion of students (1–2 in a class) who dislike speaking in front of others, in English or Japanese, would feel pressure from the introduction of a speaking exam.

Although he is not sure of how well they would score, he believes his students could cope well with the first and second BCT-S tasks because they are similar to activities conducted in class. To cope with the task demands of Tasks 3 and 4, he would give students more opportunities to discuss social problems and speak spontaneously. If the BCT-S were introduced, he imagines he would create a regular activity where students are given a social topic and one or two minutes to think about, and are then required to tell a partner. This will help students become familiar with speaking more spontaneously about a variety of social topics. Currently, he does this activity with students who come to him for additional practice for the EIKEN speaking test. Also, because most of the students in the specialized English course take the EIKEN test, he incorporates discussion of social topics in some regular classes for those groups, too. However, speaking about social topics is not something that is done for other classes taught by himself or other teachers.

Even with practice, however, discussing social topics spontaneously would be difficult for many of his students. Task 4, which requires students to plan a long response covering a number of points, would be especially challenging. He reasons that this is due to the burden the task places on working memory: having to respond to multiple points in a second



language requires a certain level of ability in the language, which many of his students simply do not have. The tablet-based format is not expected to create any difficulties as his students not only have experience of recording their voices in English classes, but they have also taken the GTEC exam.

He reports that parents are happy with the approach taken to studying English in his school, so he does not feel there would be any reaction from them.

Interview 8

Teacher 8 works at a prestigious public senior high school in an urban area in Northern Japan. The school has over 900 students most of whom will go to university. There are around ten full-time English teachers and an ALT. In line with the expectations of students and their parents, teachers at the school are heavily focused on preparation for university entrance exams:

‘They are always thinking about how many people are going to go to Hokkaido University or Kyoto, how many numbers, they are always concerned of the numbers or some of the scores of mock tests all the time.’

The number of students who gain entrance to prestigious universities is a primary goal for teachers and therefore teaching tends to follow traditional methods that are believed to support achievement of this goal (i.e., focusing on ‘teaching the mock test, grammar translations, or reading’). Teacher 8 was transferred to the school with the aim of introducing progressive methods of teaching (i.e., focusing on ‘speaking and writing’). Therefore, his views and practices of English language teaching conflict with those of the majority of teachers in the school. For instance, he wants students to learn to communicate and he worries that they will go to a top university but not be able to communicate with the international students there. Conversely, he says other teachers are only concerned with the number of students entering such institutions.

The English curriculum comprises three English Communication classes and three Logic and Expression classes, one in each grade, with an additional class that is offered in Grade 12 that focuses on presentations and speeches. For English Communication classes, the teachers in charge of each grade prepare a common handout. However, while teachers cover the same content, there is variation in methodology. Notably, Teacher 8 uses English most of the time in class while most other teachers use Japanese.

English Communication classes are generally similar across Grades 10 and 11, whereas Grade 12 classes become focused exclusively on entrance exam preparation after the summer vacation. In those classes, they ‘concentrate on taking a mock test all the time’. In a typical class, students follow the textbook and do reading and listening activities, with writing activities occupying a smaller proportion of class time. Speaking activities, including shadowing and story retelling, typically comprise up to ten minutes of a 55-minute class. However, not all teachers do this:

‘the other teachers don’t want to do [retelling], because they just don’t know what, why we are doing this or something, because they just don’t know the effectiveness of the exercise. But I always do it.’

Logic and Expression classes are quite different. There are two class hours per week, one concentrates ‘on grammatical things using some textbook in Japanese’, and the other is team-teaching with the ALT. In the former, the official textbook is not really used despite the fact that the revised textbooks have numerous activities focusing on productive skills; instead a ‘sub-textbook’ is used, which focuses on reading, grammar, and translation in preparation for prestigious national university exams. These books are used in one class in Grades 10 and 11, and for both classes in Grade 12, when entrance exams become the priority. He is disappointed with the situation and feels it is ‘boring’ and a waste of time: ‘I don’t enjoy it. But the other teachers always agree [on the curriculum plan].’ In the team-teaching classes, Teacher 8 first conducts a 15-minute speaking activity that involves students thinking about a social topic together with a partner, then the ALT will ask some questions to students and conduct a brief whole-class discussion. This is followed by a 15-minute activity focusing on a news text with listening, reading and dictation exercises. The text introduces a novel topic (e.g., whale hunting or eating dog meat) that students think about and try to convey their ideas to the ALT:

‘sometimes the students cannot explain in English, so maybe they can just explain or tell me some ideas in Japanese. Then I ask them to just use English, some sort of English to the ALT, so the ALT can understand that. And the ALT and I just talk about some like topics so that students can listen. And the students are trying to understand what we are talking about.’

The third 15-minute sequence involves speaking and writing activities, where students prepare short, organized speeches or texts. This lesson plan is conducted by Teacher 8 and the ALT a few times, after which the ALT takes the lead in delivering this lesson plan in other teachers’ classes, so they do not have to do it themselves. This is an example of how Teacher 8 is trying to incorporate speaking and writing activities into the curriculum despite the reluctance of other teachers. The first-semester Logic and Expression classes follow this speaking-centered plan, which culminates in a speech that students deliver to the class without a script, and the second semester classes focus on writing, where students write a short essay of around 150–250 words. These team-teaching classes involve a lot of listening to the teachers and some speaking in English in pairs and with the teachers.

Although teachers are expected to cover the same material in each grade, there is often a gap in the delivery of the content between Teacher 8 and other teachers, which is due to the differences in the focus and methods of teaching. Other teachers may give less time for speaking and writing activities in the class, and often they leave the teaching to the ALT and just assist in classroom management and translating when needed. As these descriptions illustrate, there are considerable differences in approaches to teaching in the school, across

both teachers and classes, but by the end of each semester students are expected to have covered the same content.

Since he joined the school, Grade 11 students have started taking the EIKEN Grade 2 Test. He and other teachers run additional after-school classes for students that advance to the second-stage speaking test. In this way, when an exam is involved, all teachers contribute to teaching speaking skills. After school, Teacher 8 coaches a small number of students who will take the EIKEN Grade Pre-1 test.

If the BCT-S were introduced, Teacher 8 would not change his overall teaching style so much. He thinks the BCT-S question types are similar to those in other speaking tests, such as EIKEN and GTEC, and therefore his teaching already covers most of the necessary skills. Nevertheless, he would expand his teaching of speaking:

‘maybe I have to spare more time practicing more in the English logic and expression class, because if the speaking test is to be introduced, you have to just prepare for it, right.’

His students would generally cope with Part 1 because simple speaking activities are done in his class. Although they do not do practice picture description much in class, students are familiar with EIKEN, which has such a task, and with the Common Test listening test, which involves listening to picture descriptions. Nevertheless, because these activities are not done in normal classes, students would need to practice more and pay attention to various aspects of the pictures. This would also help prepare students for Task 3, which involves describing two pictures, which is a task that students are not familiar with. Students would perhaps start learning how to do this from Grade 10. However, if students are still not used to speaking from their junior high school education, he would focus on getting them to speak in class during the first semester, and then introduce test-like activities (Tasks 1 and 2) from the second semester. Personal questions, such as those in Task 1, are suitable for helping learners become accustomed to speaking with a partner in class in Grade 10. Once they are used to paired speaking, they would be better prepared to do individual speeches in front of the class. Questions that focus on social topics would be introduced later, in Grade 11, once students have gotten more used to speaking in English.

A key priority would be getting students used to coming up with ideas. Teacher 8 gives the students many topics in his classes and tries to build knowledge and elicit suggestions from students. His students are not used to thinking about various social topics and coming up with ideas. While they are academically high-level, they are passive in class and ‘just don’t have any experience to express their ideas.’ This is true not only in English but also in Japanese. Students do not have enough experiences from which to draw on when responding to impromptu questions. For this reason, he emphasizes introducing various topics in class and encouraging students to read and research more about them using their iPads.

He would perhaps use textbooks that focus on news, and he would like to use foreign-made textbooks but they are expensive and there may be resistance from other teachers. Foreign

textbooks are different from the official textbooks produced in Japan because they include speaking and writing activities and using them is a form of training, so teachers can learn how to conduct production activities. Textbooks for English Communication are very limited in terms of speaking activities, he says, and in any case:

‘the other teachers always are concentrating or reading all the time. So, they just don’t ask the students to talk based on each topic or something. They just don’t want that. They don’t want to. And, as I mentioned, they’re speaking all in Japanese. So, the English teacher speaking Japanese never asks the students to talk in English, right?’

Part 4 is the most difficult and would require the most preparation. He suggests doing similar practice activities in class once or twice a month. Although students currently do speaking activities in Logic and Expression classes, some students may go off topic, may not speak for very long or at all, and may not take the activity seriously. Moreover, the goal is simply to get them to say something: ‘I force every student to just talk.’ However, all students would need to practice to the same level and they would need to concentrate on criteria assessed in the test. Currently, the criteria for assessing speaking is ‘kind of vague’ but this would need to be more detailed.

Communication and interaction ability in English and Japanese, and social skills in general, of some students is very poor. Also, roughly five students in a class are extremely quiet. He says when he was first transferred, his students couldn’t say anything to ALTs or other-language speakers who would come to the school. However, many students now try to communicate through the use of Google translation and other ‘gadgets’, whereas two years ago no-one would even try. Most of the teachers also have poor English speaking ability and ‘don’t speak English to the students’. Most of the teachers have never been abroad and are not willing to speak English in the classroom:

‘as I mentioned many times, there are only three teachers who have been abroad to get a masters or something. The other teachers always prioritize some grammar things. So even though we talk about it we just disagree with each other sometimes. So, I’m sure if those speaking tests were introduced, maybe the students, and the teachers are kind of forced to learn it. But [the teachers] would not, really, they will not, they’re not willing to introduce those kinds of things, even though they are [assessed], I think. So, I’m not sure what the speaking ability is going to be in my school. I’m sorry about it.’

Interview 9

Teacher 9 works at a prestigious public high school affiliated with a national university in an urban area of Eastern Japan. The school is relatively small with around 360 students, that is, three 40-student classes in each grade. Although it is one of the most highly ranked in terms of the difficulty to gain entrance, around half of the students enter after graduating from the affiliated junior high school, creating some disparity in terms of English levels among students. There are three full-time teachers, three part-time teachers, and an ALT.

Students take the mandatory English Communication and Logic and Expression classes in each grade, with the option to take an English Conversation class in the third year. English Communication classes follow the official textbook and are mainly based around a reading with comprehension questions and grammar point, followed by some writing and speaking activities. Logic and Expression classes are ‘grammar-oriented’ in Grade 10 but thereafter they become ‘more integrated classes and lots of writing and speaking, too’ (i.e., presentations, project work). Two classes are taught each week, one by the Japanese teacher and one with an ALT, who takes the lead in the class.

Teacher 9 often conducts a warm-up speaking-and-writing activity at the beginning of the class (in all grades) in which students each answer a question related to the topic of the reading with a partner, who then writes down the answer and submits it afterwards. This kind of speaking activity is conducted in most classes, either as a warm up or later on. Shadowing and reading aloud are also conducted in some classes. Speaking assessments are conducted in most semesters. In English Communication in Grade 11, for example, students will do a prepared one-minute speech describing a picture (‘speech test’), and group presentations in the first and second semesters, respectively. In Logic and Expression, students do, for example, a recorded interview assessment in which students interview one another (Grade 10), and a recorded debate (Grade 11). The exact type of speaking activities and assessments vary across grades and teachers. Students are also required to take the Cambridge Assessment English B1 Preliminary in Grade 10 and B2 First in Grade 11, and a few even take C1 Advanced in Grade 12.

If the BCT-S were introduced, she would do something similar to the preparation she does with students one or two months prior to them taking the Cambridge Assessment English exams (henceforth ‘Cambridge exams’). Students are familiarized with the four task formats across four classes, one task in each class. For instance, for Part 1 of the Cambridge exams, students talk about personal topics, which is similar to Task 1 in the BCT-S. In class, students work with a partner and answer questions; sometimes one student observes the pair and gives feedback. Part 2 of the BCT-S and the Cambridge exams is picture description, so in that class the teacher explains the task, gives some tips and lets students practice. Students will then answer some follow up questions and they will make some questions up by themselves to ask and answer together. Extra time would be devoted to practising the harder Tasks 3 and 4. For Task 4, she would provide brainstorming time to help students come up with ideas. She would use questions from previous test materials and begin this familiarization phase in the second semester of Grade 12. She would thus need access to a variety of past papers to use for in-class activities.

She wonders how she could give feedback to individuals, with 40 students in the class. Evaluating students’ recorded work during class would be too burdensome for the teachers and such evaluation of recordings is only done for end-of-year assessments. Evaluating written responses is one suggested way of providing feedback for Tasks 3 and 4. She would give

feedback on content, organization, use of tenses and grammar, and pronunciation if it makes speech unintelligible. She would check the BCT-S grading criteria and ‘analyse it with students’.

Teacher 9 recognises that it takes time to develop speaking skills and therefore would increase the amount of speaking done in classes, at least for Grade 12 students. However, she also recognises that the activities currently done in class are basically adequate to develop skills for the BCT-S, except for the compare/contrast in Task 3, which is unfamiliar. She would try to get students to speak more in English Communication classes ‘because right now, I know I don’t spend enough time on speaking activities.’ Some students do not like speaking in English, however, and some do not participate in such activities. This willingness to participate may be a factor in her decision about including more speaking activities. Students’ willingness to speak is dependent on class dynamics with some classes being eager to speak in English and others not. Students’ motivation to speak, however, would likely increase if they are going to be assessed as part of entrance exams:

‘Yeah, if it’s [part of the] *Kyotsu Tesuto*, then yeah, probably they would, the test would affect students’ motivation. Because everyone, all the students need to take *Kyotsu Tesuto*. So, I think it would change the atmosphere. Probably.’ (9)

Students could generally cope with the task demands if they are familiarized with them. Picture description tasks are used sometimes ‘just for fun’ in speaking activities, and there is the ‘speech test’ mentioned previously that involves describing a picture. However, it would be difficult for students to come up with ideas in Task 4. This is difficult for her students to do, in English or in Japanese, because of the lack of context. For instance, in her classes, students are introduced to topics through a reading and then asked to respond to questions on that topic, which they can do. However, in the BCT-S, test takers have no context but are expected to deliver a lengthy structured response on a new topic after only one minute of preparation time. They would therefore need more practice of formulating responses for various topics. The format would not be problematic: some students would prefer the semi-direct whereas others would predict a direct interview format.

Other teachers would likely feel similar to Teacher 9 and incorporate more speaking activities earlier on in the curriculum because they realize it takes time to develop speaking ability. All teachers are happy to speak English in class:

‘We have the similar mindset, like communicative teaching is very important. Also, like I said, some of the students are not so eager to talk, but some students love talking in English. They seem to enjoy our communicative way of teaching. They seem to think that college entrance examination-oriented teaching is boring.’

Unlike in other schools, teaching is not focused on entrance exams. Many students go to cram school to prepare for specific entrance exams and ‘they don’t expect us to teach for college entrance exam’. One reason is perhaps that the school is a designated Super Science High School, in which inquiry-based learning is expected in all subjects. Also, it may be an institutional ethos that derives from being a university-affiliated school:



'One of the teachers said the mission of high school which is attached to the university is to develop new curricular or challenging new way of teaching. So, we all feel like this pressure that we have to do something special. So, innovative teaching method or something like that.'

Students are also expected to study by themselves because 'school is not sufficient'. She tells them they must read regularly (e.g., one 400-word passage per day) and that just coming to class is not enough:

'they can study on their own without going to the cramming school. So, I don't encourage them to go to a cram school, but they need to study on their own in addition to English classes here.'

Regarding the BCT-S, some students would likely go to cram school to practice speaking if their level is not high enough. Going to speaking classes where a lot of feedback is provided would be beneficial. Not only feedback but the fact that others are there to speak to would be motivating. Practising speaking by oneself is 'not fun' because it's not 'actual communication'.