

Japanese University English Language Entrance Exams and the Washback Effect: A Systematic Review of the Research

Simon Paxton, Tae Yamazaki and Hannah Kunert*

Paxton, S., Yamazaki, T., & Kunert, H. (2022). Japanese university English language entrance exams and the washback effect: A systematic review of the research. *Journal of Pan-Pacific Association of Applied Linguistics*, 26(2), 1-20.

The entrance exams to Japanese universities have long been criticised for creating negative “washback” on English education in high schools in Japan, as teachers have often focused their teaching efforts on teaching grammar to prepare students for these exams rather than helping students to develop competency in spoken English. Many government initiatives have attempted to reform English language education by encouraging a greater focus on speaking competency. This paper outlines the nature of the Japanese university entrance exam system, considers the phenomenon of washback, and critically reviews the body of empirical research related to washback in the Japanese context. A systematic review was conducted on articles published between 1993 and June, 2022. These articles were subjected to thematic analysis in which four broad categories were identified: learner behaviour; washback to teacher behaviour; affective factors; and washback specifically from the introduction of the listening test. Findings suggest that more empirical studies are required to understand the complex nature of washback from university entrance exams, and that future research is required to examine potential washback from four-skills testing and, in particular, from the inclusion of a speaking test.

Keywords: washback, backwash, Japanese university entrance exams, communicative language teaching

1 Introduction

“Washback,” sometimes called “backwash”, refers to “the influence of tests and examinations in teaching and learning contexts” (Saville, 2009, p. 2). It is a widely discussed area in the language teaching and testing literature (see, for example, Alderson & Wall, 1993; Bailey, 1996; Brown, 1997; Hughes, 1989; Messick, 1996; Morrow, 1986, among others). Nevertheless, in Alderson and

* First/Corresponding author: Simon Paxton, Lecturer, Department of English, Faculty of Arts and Sciences, Komazawa University

Co-authors: **Tae Yamazaki**, Associate Professor, Department of English, Faculty of Arts and Sciences, Komazawa University; **Hannah Kunert**, Assistant Professor, Department of English, Faculty of Arts and Sciences, Komazawa University

Wall's (1993) seminal article, "Does washback exist?", the authors noted that the existence of washback is often based on assumptions rather than empirical evidence which they argued was lacking. Since then, however, there has been a growing body of work in this area.

The process of taking entrance exams for universities in Japan has traditionally been referred to as *shiken jigoku*, literally "examination hell", which provides a lucid, albeit somewhat outdated description of the arduous task Japanese high school students undertake when applying for entry into university in Japan. Not surprisingly, in the 1980s and 1990s, the difficulty of entrance exams for universities in Japan caught the attention of a number of researchers (see, for example, Amano, 1990; Cummings, 1980; Frost, 1991; Rohlen, 1983). While entrance exams still present a challenge for high school students trying to enter university in Japan, demographic conditions and an increasing number of students gaining entrance by recommendation have somewhat helped to lower the bar, particularly for entry into universities not considered elite institutions.

Nevertheless, English has always been an important component of university entrance exams in Japan. It is widely believed that one of the reasons Japan has not performed well in promoting English competency, particularly speaking proficiency, is due to negative washback. Criticism of English as a Foreign Language (EFL) education is often based on the assumption that university entrance exams create such negative washback which, in turn, results in poor EFL education in high schools. Because of the emphasis on English in university entrance exams, the approach to EFL education at secondary schools has often focused on the *yakudoku* method (Gorsuch, 1998; Hawkins, 2015; Hino, 1988). This word-for-word translation technique has a long history of use in preparing students for university entrance exams, and is considered useful for reading and grammar. This approach to EFL has, however, not been in alignment with government initiatives to make English language education more communicative.

Japan's Ministry of Education, Culture, Sports, Science and Technology (MEXT) has introduced a number of reforms to promote better English competency. Since 1989, MEXT has, in relation to the English taught at school in Japan, shifted its focus to "communication" and encouraged the practice of Communicative Language Teaching (CLT). The aim of CLT is to develop communicative competence, and this method of language teaching emphasises a collaborative approach (Holliday, 1994). Although earnest efforts have been made by some teachers to make English education more communicative, early research suggested that many teachers persisted with traditional forms of teaching due to the pressure to prepare students for university entrance exams (Gorsuch, 1998; Taguchi, 2005). Therefore, the high school curriculum has not always reflected MEXT's values because school teachers have been influenced by the university entrance exam requirements. This, in turn, has affected teaching methodology and assessment, as teachers

will often teach in a manner that will benefit students in tests, known as “teaching to the test” (Saito, 2019; Smith, 1991).

More recently, MEXT (2017) has proposed a plan to implement externally administered four-skills testing for university entrance exams. These tests incorporate speaking, which, in theory at least, is an incentive for students and teachers to develop English speaking proficiency. In other words, speaking tests represent an attempt to create positive washback through external motivation. Four-skills testing has, however, been met with a barrage of criticism, in relation to two main areas: the privatisation of testing; and complications surrounding inclusion of a speaking test (Abe, 2017; Haebara, 2018).

The emerging body of research on washback and the timeliness of recent governmental proposals to reform the university entrance exams system in Japan calls for a review of the literature. The purpose of this paper is to investigate the current state of research on washback in the context of Japan. Allen and Tahara (2021) reviewed washback research in the context of Japan. The purpose of their research was to facilitate future research in this area. Moreover, their review included research on tests such as the Test of English for International Communication (TOEIC). In contrast, this paper reviews and synthesises the research on washback specifically from Japanese university entrance exams. We aim to provide complex and nuanced responses to our research questions by focusing solely on research related to university entrance examinations in the context of Japan.

Furthermore, we have decided to focus on empirical studies, in line with Watanabe’s (1996) call for an increase in such research. While many theoretical articles simply state the existence of washback, the purpose of the present paper is to synthesise studies that have investigated this phenomenon empirically. In order to clarify the central issues that arise in the literature, this review employs a thematic structure. This paper is divided into three sections. Firstly, we provide a background on washback and the Japanese education system. Next, we present our research questions and methodology. Finally, we discuss our findings organised by theme.

2 Background

Since 1946, when English was first included in the curriculum, MEXT has implemented reforms to improve language education. One example of such a reform was the introduction of an English listening test as a part of the National Center Test for University Admission (NCT) in 2006. The NCT is a standardised test, which is administered every year and used for admission to public and some private universities. It is a high-stakes test, because its scores may determine whether students are allowed to take a university administered entrance exam which follows the NCT. The listening test was added to the

written test because MEXT aimed to foster students' communication ability. Another major step by MEXT (2003) was its announcement of its 2003 action plan to Cultivate "Japanese with English Abilities". This plan emphasized communicative competence and introduced a number of measures including improving the qualifications of English teachers and promoting overseas study among upper secondary school students.

MEXT's emphasis on communicative competence is significant and requires clarification. Communicative competence can include a broad spectrum of competencies, including linguistic, discourse, socio-linguistic, and strategic competence (Ellis, 2015). The meaning of this term therefore is open to a variety of interpretations. However, whereas the Japanese educational system had previously focused heavily on English grammar and students were largely passive participants in the classroom, MEXT's use of the term "communicative competence" represented a shift to a style of English education in which students were active participants, and one in which classroom activities were conducted in English. The desired outcome is that students develop greater oral communication skills in English.

More recently, as mentioned previously, MEXT proposed to implement four-skills testing, which is another example of efforts to create positive washback by "improving the exam". According to MEXT (2017), entrance exams will be reformed with the introduction of new four-skills tests to be provided by external organisations. These reforms are being implemented in order to align university entrance exams with the national course of study which emphasises communicative competence of English.

While it is generally accepted that washback does exist, it is a multifaceted phenomenon, and Alderson and Wall (1993) criticised the washback hypothesis as being "unduly simplistic" and for making too many assumptions about how people are influenced by tests. There are many other factors, besides the exam, that can also influence washback. Both Alderson and Wall (1993) and Spratt (2005) note that teacher competence, school resources and teachers' understanding of tests can all influence teaching. Watanabe (1996) also emphasises the problematic nature of washback and suggests that research on washback should be conducted from multiple perspectives.

Moreover, washback can be either beneficial or harmful (sometimes called "positive washback" and "negative washback", respectively). Generally speaking, positive washback is defined as a positive influence of a test on teaching and learning, and negative washback is defined as the negative influence of a test on teaching and learning. The distinction between positive and negative washback, however, is not as clear as one might expect. For example, if a test functions as a form of extrinsic motivation, even a poor test can, by way of influencing learners to do more work, produce positive washback. That is, of course, if "doing more work" can be determined to be a positive outcome. Similarly, tests, whether good or poor, may produce negative washback as a result of the anxiety that they could potentially cause

students and teachers. However, some key features of both positive and negative washback outlined by Pan (2009), are consistent with both the findings of our review, and our commonsense notions of what a positive or negative influence may be. Positive washback has been achieved when teachers are encouraged to teach their subjects more thoroughly, while negative washback involves narrowing the curriculum and “teaching to the test”. From the students’ point of view, positive washback involves having a sense of accomplishment, and learning a wide range of usable knowledge, while negative washback involves cramming discrete points of testable knowledge, and accompanying feelings of anxiety.

While university entrance exams in Japan have long been blamed for poor levels of spoken English proficiency among Japan’s population, some scholars have disputed the validity of these criticisms, claiming that the exam-pedagogy relationship has been used as a convenient scapegoat (Guest, 2000; Mulvey, 2001; Stout, 2003; Watanabe, 1996). Mulvey (2001) was quite adamant that university entrance exams were being unduly blamed. Stout (2003) concurred with Mulvey, stating that the negative influence of university entrance exams had subsided, but thought that there was hope for promoting positive washback.

3 Research Questions

The central focus and purpose of this study is to investigate the state of research on washback from university entrance exams in Japan. While it is generally assumed that washback in the context of Japan does exist, a review of the literature is needed to clarify this assumption. This research interest evolved from a much broader interest in how Japanese university entrance exams can be employed to promote positive washback in English education in Japan. The following research questions are thus formulated:

RQ1: Do empirical studies on washback from Japanese university exams provide convincing evidence for this phenomenon?

RQ2: What have been the significant findings in studies conducted on washback from Japanese university exams?

4 Method

This study utilises a systematic review as its methodology. As a transparent methodology, this makes the study “reproducible and updateable” (Zawacki-Richter et al., 2019, p. vi), and allows readers to “read about relevant findings in condensed form and subsequently act upon them” (Macaro, 2019, p. 231). Macaro (2019) notes that systematic reviews can also consolidate research

evidence on an “emerging phenomenon”, making it an appropriate methodology for the evolving literature on the phenomenon of washback.

This systematic review was conducted by three reviewers, in line with Macaro’s (2019) suggestion that systematic reviews should be carried out by a team of at least two reviewers so that varying perspectives can be applied to the review. The search was conducted over June 10–13, 2022, and covered research conducted since the publication of Alderson and Wall’s (1993) seminal article.

Our inclusion criteria were that the studies:

1. are about washback from exams related to university entry in Japan;
2. contain empirical data;
3. are published between 1993—June 2022;
4. are an academic article or thesis;
5. have washback as one of the major focuses;
6. have the full text available online without payment.

The following databases were searched to find potential articles:

1. ERIC
2. Google Scholar
3. EBSCOhost

In each of the databases, the following English and Japanese keywords were searched within our designated publication years:

1. Washback AND Japan
2. Backwash AND Japan
3. Washback AND Japanese university entrance examinations
4. Backwash AND Japanese university entrance examinations
5. 波及効果 大学入試 英語
(Washback effect, university entrance examinations, English)
6. ウォッシュユバック 大学入試 英語
(Washback, university entrance examinations, English)
7. バックウォッシュユ 大学入試 英語
(Backwash, university entrance examinations, English)

Results were imported into Rayyan, an online research tool for conducting systematic reviews. The initial search produced 58 potential papers. Each article was then independently assessed by reviewers for possible inclusion or exclusion by applying the criteria stated above. The reference lists of those papers were also examined to determine whether any relevant literature appeared that was not produced by the database searches.

The resulting 26 articles were assessed independently by the three authors to identify themes relevant to the research questions, and were labelled under the “blind” function within the Rayyan program. Previous discussions

had identified some expected themes, such as washback on student or teacher behaviour. In addition, each author also added new themes as they arose from the data, such as “washback specific to the listening exams”. Then, the blind function was turned off, and labels across the three authors were compared. A high level of consistency was revealed, and minor differences were resolved by discussion. Some themes were collapsed and others broken down into separate categories, in order to give the clearest picture of the collected findings. The final thematic labels and the relevant studies are presented in Table 1 below. Most scholarship on washback in the context of Japan has focused predominantly on washback to learner behaviours and washback to teacher behaviours, which were two major themes identified in the reviewed studies. However, our review uncovered other themes including affective behaviours (beliefs, motivation, etc.) and washback specifically from the introduction of the listening test in 2006. As can be seen in Table 1, in some cases, multiple themes arose in the same paper.

Table 1. Results of Systematic Review

Themes	Studies
Washback to learner behaviours	Allen (2016a); Allen (2016b); Bailey (2018); O’Donnell (2003); Otaka (2011); Sato (2019); Takagi (2010); Yanagawa (2012); Kowata (2009); Oguri (2009)
Washback to teacher behaviours	Bailey (2018); Caine (2005); Kowata (2009, 2015); Oguri (2009) Otaka (2011); Ozaki (2010); Shiokawa (2018a, 2018b); Taguchi (2005); Takagi (2010); Tsushima (2011); Yanagawa (2012); Shea (2009)
Affective factors (attitude, motivation, etc.)	Allen (2016b); Baba (2019); Bailey (2018); Caine (2005); Hirai et al. (2013); Kowata (2009, 2015); Nagatomo & Allen (2019); O’Donnell (2003); Oguri (2009); Otaka (2011); Sato & Ikeda (2015); Sato (2019); Taguchi (2005); Takagi (2010); Tsushima (2011); Yanagawa (2012)
Washback from the listening test introduced in 2006	Hirai et al. (2013); Oguri (2009); Saito (2019); Takeuchi & Kozuka (2010); Uchida et al. (2018); Yanagawa (2012)

5 Results and Discussion

This section presents the findings of our review of washback in relation to Japanese university entrance exams to answer the above two research questions. Firstly, we answer RQ1 by investigating whether the studies reviewed provide sufficient evidence of washback occurring. Next, we address RQ2 and report on the significant findings from the research.

5.1 Evidence of washback

This section responds to RQ1, which aims to investigate evidence of washback in the studies reviewed. As noted in Section 1, washback is often presumed to exist, but we explore whether there is sufficient evidence to prove its occurrence in the context of Japan.

Many studies showed evidence of washback from the entrance exams. In other words, students and teachers engaged in behaviours or experienced feelings that would not otherwise reasonably be expected to occur had it not been for the upcoming exams. Such behaviours include studying previous or mock exam papers, doing translation or *yakudoku* exercises, memorising complex grammar patterns, and attending *juku*, “cram” schools in Japan where students often go to undertake exam preparation courses (eg. Allen, 2016a; Nagatomo & Allen, 2019; O’Donnell, 2003; Sato, 2019; Taguchi, 2005; Takagi, 2010). As mentioned earlier, defining these behaviours as negative or positive washback depends on the wildly divergent goals of the entrance exams: entering a university; or increasing communicative competence. This mismatch, between the stated goals of MEXT and the washback that the current entrance exams appear to be generating, indicates that significant changes need to occur in the current system (Tsushima, 2011).

By contrast, some studies conclude that washback does not occur or only occurs in some areas under investigation. Caine (2005), for example, observed washback at the level of content but not at the level of methodology. Similarly, Tsushima (2011) found a lack of washback in regard to teaching practice, although washback was observed in the assessment of the class under observation. These two studies are discussed further in Section 5.2.2.

Three papers in our review, Hirai et al. (2013), Takeuchi and Kozuka (2010) and Yanagawa (2012), specifically investigated the implementation of the listening component of the NCT, which was hoped to create positive washback in terms of encouraging students to improve their listening skills. Both Hirai et al. (2013) and Takeuchi and Kozuka (2010) found, at best, limited evidence of washback when comparing student listening scores. Yanagawa’s (2012) results were even less convincing, leading him to conclude that the exam has “a limited effect (no washback)” as students spent very little time either in class or outside of class listening to English. These studies relating to the listening test are discussed in more detail in Section 5.2.4.

5.2 Thematic findings

In this second section of our results and discussion, we seek to understand the major contributions from the studies reviewed. We examine the research according to the major themes identified.

5.2.1 Washback to learner behaviours

Many studies focused on how students prepared for the entrance exams, providing evidence of washback (Allen, 2016a; Bailey, 2018; Kowata, 2015; Nagatomo & Allen, 2019; O'Donnell, 2003; Oguri, 2009; Sato, 2019; Takagi, 2010; Yanagawa, 2012). In other words, students do many things specifically to prepare for the exams, both inside and outside of school. These findings were gained from questionnaires and interviews with students, ranging from multiple rounds of interviews with a small number of students (Nagatomo & Allen, 2019) to a questionnaire administered to nearly 400 students (Yanagawa, 2012). Some general traits of student behaviour related to exam preparation can be described. These can be divided into: where students study; what they study; and some final comments on the role of student agency in managing their study.

Students' preparation was found to occur in three specific contexts: at school, at *juku*, and by themselves. To begin with, many respondents described preparing for entrance exams at school (Nagatomo & Allen, 2019; O'Donnell, 2003; Otaka, 2011; Sato, 2019; Takagi, 2010), which suggests that washback to the school curriculum is occurring. Although Otaka (2011) reports that both middle and high schools teach mostly "exam English", Sato's (2019) interviewees also mentioned a range of classroom activities not directly related to exam preparation, including shadowing, reading newspapers, and watching TED talks. In addition, Nagatomo and Allen's (2019) interviewees suggested that, in the second year of high school at least, classes covered all the four core skills of reading, writing, speaking and listening, although, as mentioned earlier, the time devoted to exam preparation increased as the exam period drew closer. Takagi's (2010) respondents reported a school syllabus weighted heavily towards exam preparation, with additional classes on Saturdays which, although voluntary, 80% of students attended. It therefore appears that, although exam preparation activities were widely reported to occur in school, the amount of time devoted to them varied by school and year level. O'Donnell's (2003) questionnaire is significant in revealing the mixed emotions elicited by the exam preparation classes: they were said to be unpleasant, complicated, and not enjoyable, but also to be rewarding and important.

Many respondents to the surveys and interviews also reported preparing for exams at *juku* (Allen, 2016a; Nagatomo & Allen, 2019; O'Donnell, 2003; Sato, 2019; Takagi, 2010). Unlike high school classes, the study at *juku* is almost always exclusively focused on exam preparation, which strongly suggests that washback is occurring to student behaviour outside of school, as it seems clear that students would not be attending *juku* if it were not for the upcoming entrance exams. Allen (2016a) describes a typical lesson sequence in which students completed part of a previous exam within a time limit and submitted their papers to the teacher. The teacher would then explain the answers in Japanese, in a teacher-centred or lecture style format, and distribute

scores the following week. *Juku* classes mostly focused on reading and grammar (Allen, 2016a; Nagatomo & Allen, 2019; O'Donnell, 2003).

Finally, the data reveal that many students prepare by themselves (Allen, 2016a; Sato, 2019; Takagi, 2010), which is more evidence of washback to student behaviour. In particular, Sato's (2019) data highlight the self-directed nature of this aspect of study. Many of his interviewees reported preparing for more than one exam (for example, for multiple universities and sometimes also the NCT or Test of English for Academic Purposes, "TEAP"), and therefore prioritising skills covered in more than one exam or than those needed for the exam of their first choice of university. Students also adjusted their study according to their strengths and weaknesses and sections that were perceived to be more or less difficult. Sato also found that students could be quite critical of advice from teachers and other sources and would employ it only when they thought the strategies were effective and useful. While much of the literature on washback suggests a clear-cut cause and effect relationship, Sato's data are thus useful in highlighting the autonomy of learners as active agents in this process.

As for what students study, there is overwhelming consistency across the articles in the review and between the study at school, *juku*, or individually. The use of past or mock exam papers was common (Allen, 2016a; Nagatomo & Allen, 2019; Oguri, 2009; Sato, 2019; Takagi, 2010), as was translation or *yakudoku* (Nagatomo & Allen, 2019; O'Donnell, 2003; Sato, 2019; Takagi, 2010), both prototypical exam preparation strategies. Vocabulary exercises were also widely reported (Allen, 2016a; Nagatomo & Allen, 2019; Oguri, 2009; Takagi, 2010; Yanagawa, 2012), the popularity of which may stem from their application across multiple sections of the exams, for example reading, writing and vocabulary. Other reported activities include writing exercises (Allen 2016a; Sato 2019) and the use of textbooks specific to exam preparation (Nagatomo & Allen, 2019; Oguri, 2009).

A final comment on washback to learner behaviour, briefly touched upon above, relates to the degree of agency students experience in preparing for exams. While it is true that the entrance exams place a great amount of pressure on students and that many adapt their study habits to increase their chances of passing, there is much individual difference in how this occurs and the extent to which learners are satisfied with the current situation. Bailey's (2018) respondents, for example, reported that speaking English was more important than studying for the exams, but also that they spent more time studying for the exams than on developing their communicative language skills. This is strong evidence of negative washback occurring, in that the exam is leading these students to behaviours that go against not only their beliefs about the importance of communicative English but also the goals of MEXT and CLT.

Some of Oguri's (2009) respondents also expressed a wish for more communicative English classes, for example opportunities to speak with the

school's Assistant Language Teacher (ALT). Sato (2019) points out that the degree to which positive or negative washback occurs is often related to a student's ability to correctly identify the relevant KSAs (Knowledge, Skills and Abilities) for the exam. This means that the same exam could generate either kind of washback; for example, a student who perceives the necessary KSAs as involving doing endless mock exams (negative washback), in contrast to a student who interprets the KSAs as including natural as well as textbook English and listening to English music and sending emails to friends in English (positive washback). It is therefore necessary to avoid making generalisations about the cause-and-effect nature of washback and to remind ourselves that students, as well as teachers, are active stakeholders in this process.

5.2.2 Washback to teacher behaviours

This section explores the research in relation to washback on teaching behaviours. Our review found a number of studies indicating that exams promoted negative washback in this area (Bailey, 2018; Oguri, 2009; Otaka, 2011; Ozaki, 2010; Taguchi, 2005; Takagi, 2010).

Bailey's (2018) findings on teacher behaviours suggest that washback from the grammar-based entrance exams deter teachers from developing their students' communicative abilities. He also found that there is a mismatch between the positive attitudes held by teachers towards CLT and the extent to which they used it in their classrooms. While the majority of the teachers thought test preparation and development of communicative ability were equally important, the time allocated to communicative activities decreased between second and third year. This indicates that teachers are more likely to be affected by the entrance exams as the exams draw nearer. The teachers believed that preparing their students for entrance exams was very important, so they would teach what is relevant to the exams even though they have a desire to teach more communicative English.

Several studies refer to practices of high school teachers to mitigate negative washback (Caine, 2005; Taguchi, 2005; Tsushima, 2011). Caine (2005), for example, investigated the effects of multiple existing English tests (i.e., university entrance exams, regional tests, and internal end of term tests) on teaching pedagogy, and also proposes an original, direct test of speaking. Classroom observation and teacher and student questionnaires were used to examine the washback effects of the current range of English tests being taken by the learners, and the new speaking test. He suggests that changes to exams from grammar-based to communicative ones would not necessarily cause positive changes to teaching methodology, as his observations suggest that grammar translation methodology was used even in classes for students not intending to take entrance exams. Furthermore, results from the questionnaire indicate that, although all the teachers showed positive attitudes to the novel speaking test, they would not change their teaching methodology. Rather than washback, it might be that teachers simply prefer a teacher-led, grammar-based

approach; and Caine therefore suggests that changes to the entrance exams should be brought about with a revision to teacher training programmes, enabling teachers to familiarise themselves with the basic principles of CLT.

Tsushima (2011) investigated washback from university entrance exams to high school teachers' perception of speaking-focused courses called "Oral Communication" (OC). Based on data collected through a questionnaire, classroom observations, an analysis of internal exams, and interviews, she argues that university entrance exams have a strong negative impact on classroom practice. Specifically, the data indicate that OC courses are taught in teacher-centred, grammar-oriented and monolingual settings, implying that teachers are mainly concerned with preparing their students for entrance exams. Like Caine (2005), she points out the significance of "a teacher factor" in washback with regard to a lack of CLT at schools. Qualitative results reveal that teachers don't have enough confidence in their own speaking abilities, which possibly discourages them from properly assessing their students' speaking abilities as well as in teaching speaking skills. Furthermore, in line with Caine (2005), Tsushima (2011) argues that teachers should be given more opportunities to update their knowledge and skills so that they become able to implement appropriate classroom teaching and testing practice in CLT settings.

In summary, several studies have identified negative washback from university entrance exams to teaching pedagogy. Furthermore, the studies reviewed suggest that changing entrance exams in order to test communication abilities may not necessarily result in positive washback, because many teachers are unlikely to implement CLT in their classrooms.

5.2.3 Affective factors

One aspect of washback that has interested researchers is the influence of university entrance exams on teachers' and students' motivation, beliefs, perceptions and attitudes. Because such factors relate to emotions and feelings, we used the term "affective" to identify this theme in our systematic review. Over fifteen papers included in our systematic review considered affective factors to some degree.

Despite the Japanese government's efforts to encourage teachers to adopt CLT, as mentioned earlier, many teachers feel pressure to "teach to the test". Taguchi (2005) investigated teachers' perceptions of oral communication classes and difficulties related to their implementation. She found that teachers are under pressure because both students and their parents expect English teachers to teach for the exams. In addition, teachers feel pressure from their school administrators who expect teachers to achieve their schools' goals and to improve their reputations by helping as many students as possible pass university exams.

This is supported by Takagi (2010) who investigated the impact of university entrance exams on the perceptions and attitudes of students and teachers toward English exams. She found that university entrance exams had

a negative impact on both language learning for students and high school teaching for teachers. Moreover, she found that the students felt stress while preparing for exams, and teachers felt pressure from students, parents and administrators. Thus, teachers appear to ignore the guidelines set forth by MEXT which emphasise communicative ability, and instead concentrate on teaching for exams.

As mentioned in Section 5.2.2, Caine (2005) found that “teacher perceptions of communicative language teaching are in many ways influenced by the existence of a hidden syllabus” (p. 12). That is, university entrance exams shape the way teachers feel about and approach CLT. Tsushima (2011), also addressed in Section 5.2.2, found that a lack of confidence in assessing students’ speaking due to their anxiety as non-native English speakers emerged as an influential factor that hinders teachers from implementing the course objectives.

Sato and Ikeda (2015) investigated test-taker perceptions of the skills being measured by high-stakes exams administered in Japan and South Korea. They found that some types of test items potentially induce a gap between test committee intentions and test-taker perceptions, which does not necessarily result in the intended washback on students’ learning.

Learners’ perceptions can also relate to a number of different areas. For example, while perceptions of the abilities being tested in high-stakes tests can clearly affect learner behaviour, learners’ perceptions of the effectiveness of certain learning activities may influence which strategies learners decide to implement (Allen, 2016a). Nagatomo and Allen (2019) looked at beliefs about English and their study methods. They found that many students found studying for entrance exams to be demotivating, although they acknowledged the importance of the process and were aware of many of the benefits of learning English.

Baba (2019) investigated learners’ beliefs from the perspective of educational psychology and considered the kinds of beliefs that lead to creating beneficial washback. For example, drawing on the work of psychologists, Baba notes how the competitive nature of entrance exams can be perceived as a source of stress or an opportunity to grow. As one would expect, he notes how the perception of entrance exams as an opportunity to grow is a more positive attitude which is reflected in students’ approach to preparing for tests and in general results in higher motivation and autonomous learning.

The importance of attitude in language acquisition is well-established. O’Donnell (2003) investigated the attitudes that first-year university students held towards language learning, and found that most students had a negative attitude towards their experiences at high school. He concludes that language teachers at university should be aware of their students’ prior experiences and their consequent motivation and attitudes held towards language learning.

In summary, an investigation of the literature on washback in relation to affective factors highlights the importance of researching this area.

Importantly, many studies note the psychological impact that exams have on students and the extreme pressure teachers are under to prepare students for entrance exams. Finally, despite efforts to promote positive washback by changing tests or the curriculum, such efforts may be in vain if they are not supported by similar changes in students' and teachers' beliefs, perceptions, attitudes and motivation. As Chapman and Snyder (2000) suggest, it is not so much changes made to tests that generate washback, but teachers' beliefs about those changes.

5.2.4 Washback from the introduction of a listening test

In this section, we review studies that address washback from the introduction of the listening test.

Our review found studies that examined whether or not the introduction of the listening test may improve students' listening ability. Takeuchi and Kozuka (2010) compared TOEIC listening scores of a national university's first-year students across majors for four years (2005-2008). All the students were required to take the NCT for admission to the university. They found that the introduction of the listening test to the NCT did not improve the students' listening skills across the board; namely, students who majored in international cultures showed a statistically significant increase in their listening scores after 2006, but those who majored in early childhood education did not. This suggests that positive washback from the introduction of the listening test would be expected only when students are willing to enhance their communication ability after they enter university.

Similarly, Hirai et al. (2013) examined changes in first-year university students' listening scores on the placement test administered by the university for 10 years (2002-2011). They found that the students' listening skills did not necessarily improve after the listening test was introduced and that the positive impact from the introduction of the test was limited to students in a humanities department who majored in languages and cultures. It is thus concluded that the introduction of the listening test might give a relatively stronger impact on students who are highly motivated to acquire communication skills in English. Hirai et al. further suggest that the NCT listening test needs adjustment so that MEXT's goal to develop students' communication ability is achieved.

Further studies have been found that have examined washback to learners' and teachers' behaviours. Opinions vary as to whether positive washback was observed. For instance, Hirai et al. (2013) additionally investigated students' behaviours to prepare for the listening test and their attitudes toward the inclusion of the listening test in the NCT. Responses from first-year students who had taken the NCT listening test reveal that they did not spend much time preparing for the listening test, suggesting that the test did not strongly motivate the students to improve their listening ability. However, most of the students favoured the listening test introduced to the NCT, because they realised that improving listening ability is important for

them to be able to communicate in English as well as to pass the exams. This suggests that the introduction of the listening test may have had a positive impact on students' attitudes toward studying listening.

Yanagawa (2012), by contrast, found that the introduction of the test had limited influence on both learner behaviours and teaching pedagogy. He administered a questionnaire to high school teachers and students about whether they are interested in the NCT listening test and how much time they spent teaching or learning listening skills, respectively. Their responses reveal that, while the vast majority of the teachers had an interest in the NCT listening test, most of them spent less than 30% of class time on listening instruction. Likewise, the majority of the students were interested in the test but spent little time studying listening, partly because they did not know how to do such study. Therefore, he concludes that the introduction of the NCT listening test has not created positive washback.

To sum up, our review of studies focusing on washback from the introduction of the listening test to the NCT has found that, although it did not contribute to enhancing Japanese learners' listening ability, it may help them to have more positive attitudes toward studying listening.

6 Conclusion

Given the widespread assumption that entrance exams for universities in Japan create negative washback, as well as the recent government proposals to introduce four-skills testing for university entrance exams, this review aimed to consolidate the relevant empirical research conducted in this important area. Our intention in this review was not to judge EFL education in Japan but to investigate what research has been undertaken and where future research may lead. The results from our review, however, highlight critical issues about washback in the context of Japan.

Based on the review of the literature and in response to RQ1, it is apparent that there is evidence of washback. However, the occurrence of washback is not straightforward. Ultimately, as Sato (2019) notes, students are autonomous learners and make decisions about what to study. Furthermore, whether washback is positive or negative depends in part on students' ability to correctly identify key areas to study for entrance exams. Therefore, determining whether washback is positive or negative is, in many cases, a subjective, or at least complex judgement. O'Donnell (2003), for example, notes that some students believe that studying for exams improves their general proficiency in English. While there may be some truth to this, this type of study is unlikely to be conducive to improving communicative English skills, particularly speaking competency. In any case, our findings are consistent with Alderson and Walls' (1993) assertion that washback is a complex phenomenon.

In terms of significant findings from the research reviewed, a number

of interesting discoveries were made. Firstly, the research demonstrates how the washback effect on learner behaviour is not definitive, because ultimately students have individual differences. Secondly, changing entrance exams to test communication abilities does not guarantee positive washback, because teachers still may not implement CLT. Further teacher training is needed for teachers to adopt CLT in their classrooms. Moreover, any change to exams must be supported by a change in beliefs, attitudes and motivation. Finally, research on washback from the listening test reveals that, even though the test did not seem to contribute significantly to enhancing listening ability, it might have helped in fostering more positive attitudes toward learning listening. This is significant as one might consider how the implementation of a speaking test could result in a similar outcome.

Overall, recent scholarship in this area, in addition to the potential for future government initiatives, suggests that washback will attract more attention from scholars in the future. Indeed, longitudinal research is needed to provide more empirical research on this important phenomenon. In the future we hope to see more studies emerge to address washback from the introduction of four-skills testing. However, just as our review reveals that the introduction of a listening test did not have a major influence on teaching pedagogy, it is difficult to foresee whether the inclusion of a speaking test will have an impact. Of course, devising a suitable speaking test that can measure communicative competence, especially oral communication skills, also presents several challenges (Saito, 2019), and there are several valid criticisms of four-skills testing. Nevertheless, the inclusion of a speaking test might encourage students and teachers to focus more on speaking, which may bring English education at high school more in line with MEXT's goals of making English education more communicative.

References

- Abe, M. (2017). *Shijō saiaku no eigo seisaku—uso darake no “4 ginō” kanban*. [English education in chaos: Confusion and dishonesty in Japanese government policy]. Hitsuji Shobō.
- Alderson, J., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14, 115-129.
- Allen, D. (2016a). Japanese cram schools and entrance exam washback. *The Asian Journal of Applied Linguistics*, 3(1), 54-67. Retrieved from <https://caes.hku.hk/ajal/index.php/ajal/article/view/338>.
- Allen, D. (2016b). Investigating washback to the learner from the IELTS test in the Japanese tertiary context. *Language Testing in Asia*, 6(1), 1-20.
- Allen, D., & Tahara, T. (2021). A review of washback research in Japan. *JLTA Journal*, 24, 3-22.

Japanese University English Language Entrance Exams and the Washback Effect: A Systematic Review of the Research

- Amano, I. (1990). *Education and examination in modern Japan*. Tokyo University Press.
- Baba, S. (2019). Dō sureba haisuteikusu tesuto no nozomashii hakyū kōka wo motarasu koto ga dekiru no ka? Kyōiku shinrigaku kenkyū kara no teian [How to Produce Beneficial Washback Effect by Using High-Stakes Testing? Proposal From Educational Psychology]. *JLTA Journal*, 22, 44-64.
- Bailey, J. L. (2018). A study of the washback effects of university entrance examinations on teaching pedagogy and student learning behaviour in Japanese high schools. *British Journal of Education*, 6(6), 50-72.
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257-279.
- Brown, J. D. (1997). Do tests washback on the language classroom? *TESOL ANZ Journal*, 5, 63-80.
- Caine, N. (2005). *EFL examination washback in Japan: Investigating the effects of oral assessment on teaching and learning* [Unpublished Master's thesis]. University of Manchester.
- Chapman, D. W., & Snyder Jr, C. W. (2000). Can high stakes national testing improve instruction: Reexamining conventional wisdom. *International Journal of Educational Development*, 20, 457-474.
- Cummings, W. K. (1980). *Education and equality in Japan*. Princeton University Press.
- Ellis, R. (2015). *Understanding second language acquisition*. (2nd ed.). Oxford University Press.
- Frost, P. (1991). Examination hell. In E.R. Beauchamp (Ed.), *Windows on Japanese education* (pp. 291-305). Greenwood Press.
- Gorsuch, G. J. (1998). Yakudoku EFL instruction in two Japanese high school classrooms: An exploratory study. *JALT Journal*, 20(1), 6-32.
- Guest, M. (2000). "But I have to teach grammar!": An analysis of the role "grammar" plays in Japanese university English entrance examinations. *The Language Teacher*, 24(11), 23-29.
- Haebara, T. (Ed.). (2018). *Kenshō meisō suru eigo nyūshi supīkingu dōnyū to minkan itaku*. [Investigation: A wandering debate over English entrance exams – The introduction of a speaking test and the implementation of externally-administered tests]. Iwanami Shoten.
- Hawkins, S. J. (2015). Guilt, missed opportunities, and false role models: A look at perceptions and use of the first language in English teaching in Japan. *JALT Journal*, 37(1), 29-42.
- Hino, N. (1988). Yakudoku: Japan's dominant tradition in foreign language learning. *JALT Journal*, 10(1), 45-55.
- Hirai, A., Fujita, R., Ito, M., & O'ki, T. (2013). Washback of the center listening test on learners' listening skills and attitudes. *ARELE: Annual Review of English Language Education in Japan*, 24, 31-45.

- Holliday, A. (1994). The house of TESEP and the communicative approach: The special needs of state English language education. *ELT Journal*, 48(1), 3-11.
- Hughes, A. (1989). *Testing for language teachers*. Cambridge University Press.
- Kowata, T. (2009). Juken-sha no raitingu mondai no toraekata to sono taisaku - nihon no daigaku nyūgaku shiken kankyō ni okeru hakyū kōka [How writing tests interpreted and prepared for? Washback effect in the context of Japanese university entrance examinations]. *ARCLE Review*, 3, 122-132.
- Kowata, T. (2015). *Washback effects of university entrance examination writing tasks on learning and teaching* [Unpublished doctoral dissertation]. Tokyo University of Foreign Studies.
- Macaro, E. (2019). Systematic reviews in applied linguistics research. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics* (pp. 230-239). Routledge.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241-256.
- MEXT. (2003). Action plan to cultivate “Japanese with English abilities”.
- MEXT. (2017). Daigaku nyūgakusha senbatsu kaikaku ni tsuite [Reforms to the Selection of University Entrants]. Retrieved from http://www.mext.go.jp/b_menu/houdou/29/07/1388131.htm.
- Morrow, K. (1986). The evaluation of tests of communicative performance. In M. Portal (Ed.), *Innovations in language testing* (pp. 1-13). NFER/Nelson.
- Mulvey, B. (2001). The role and influence of Japan’s university entrance exams: A reassessment. *The Language Teacher-Kyoto-JALT*, 25(7), 11-17.
- Nagatomo, D., & Allen, D. (2019). Investing in their futures: Highly-motivated students’ perceptions of TEAP and university entrance exams. *The Language Teacher*, 43(5), 3-7.
- O’Donnell, K. (2003). Uncovering first year students’ language learning experiences, attitudes, and motivations in a context of change at the tertiary level of education. *JALT Journal*, 25(1), 31-62.
- Oguri, Y. (2009). Washback effects from the listening section of the national center test for university admissions: Suggestions for high school classes. *Academic Reports of the University Center for Intercultural Education, The University of Shiga Prefecture*, 14, 35-44.
- Otaka, H. (2011). The backwash effect of the entrance exam on English language education in Japan. *Gengo bunka ronshū*, 7, 161-184.
- Ozaki, S. (2010). High-Stakes tests and educational language policy in Japan: An analysis of the National Center English Test from the viewpoint of education for international understanding. *TESOL Journal*, 2, 175-188.
- Pan, Y. C. (2009). A review of washback and its pedagogical implications. *VNU Journal of Foreign Studies*, 25(4), 257-263.

Japanese University English Language Entrance Exams and the Washback Effect: A Systematic Review of the Research

- Rohlen, T. (1983). *Japan's high schools*. University of California Press.
- Saito, Y. (2019). Impacts of introducing four-skill English tests into university entrance examinations. *The Language Teacher*, 43(2), 9-14.
- 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.
- Sato, T., & Ikeda, N. (2015). Test-taker perception of what test items measure: A potential impact of face validity on student learning. *Language Testing in Asia*, 5(10), 1-16.
- Saville, N. (2009). *Developing a model for investigating the impact of language assessment within educational contexts by a public examination provider* [Unpublished doctoral dissertation]. University of Bedfordshire.
- Shea, D. P. (2009). The university English entrance exam and its implied effect on EFL pedagogy in Japan. *The Hiyoshi Review of English*, 55, 97-132.
- Shiokawa, H. (2018a). Daigaku nyūshi jiyū eisakubun rondai ni okeru janru (tekisuto taipu) no tayōsei no tankyū [Exploring the variety of genres in writing tasks in Japanese university entrance examinations.] *Kantō kōshinetsu eigo kyōiku gakkaiishi*, 32, 15-28.
- Shiokawa, H. (2018b). Daigaku nyūshi jiyū eisakubun rondai janru bunseki—kako 17 nenkan no kokuritsu daigaku no nyūshi mondai dēta ni motozuite [Genre analysis of writing tasks in Japanese university entrance examinations: Based on the data collected from entrance exams of national and public universities.] *Chūbu chiku eigo kyōiku gakkai kiyō*, 47, 49-56.
- Smith, M. L. (1991). Meanings of test preparation. *American Educational Research Journal*, 28(3), 521-542.
- Spratt, M. (2005). Washback and the classroom: The implications for teaching and learning of studies of washback from exams. *Language Teaching Research*, 9(1), 5-29.
- Stout, M. (2003). Not guilty as charged: Do the university entrance exams in Japan affect what is taught? *The ETJ Journal*, 4, 1-7.
- Taguchi, N. (2005). The communicative approach in Japanese secondary schools: Teachers' perceptions and practice. *The Language Teacher*, 29(3), 3-12.
- Takagi, A. (2010). *A critical analysis of English language entrance examinations at Japanese universities* [Unpublished doctoral dissertation]. University of Exeter.
- Takeuchi, T., & Kozuka, Y. (2010). An analysis of TOEIC scores before and after the introduction of listening test of the national center for university entrance examinations: A case study of international cultural studies majors, English majors and early childhood majors. *Aichi Kyōiku Daigaku Kyōiku Jissen Sōgō Center Kiyō*, 13, 127-131.

- Tsushima, R. (2011). *The mismatch between educational policy and classroom practice: EFL teachers' perspectives on washback in Japan* [Unpublished master's thesis]. McGill University.
- Uchida, T., Sugisawa, T., & Kei, I. (2010). Structure of academic skills measured by the listening comprehension test in National Center Test, *Japanese Journal for Research on Testing*, 6(1), 103-111.
- Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing*, 13, 318-333.
- Yanagawa, K. (2012). Do high school English teachers teach listening, and do students listen to spoken English? —To explore washback effects of the introduction of the Centre listening test in Japan. *Dialogue*, 11, 1-14.
- Zawacki-Richter, O., Kerres, M., Bedenlier, S., Bond, M., & Buntins, K. (Eds.). (2019). *Systematic reviews in educational research: Methodology, perspectives and application*. Springer VS.

Simon Paxton, Lecturer
Komazawa University
1-23-1 Komazawa, Setagaya-ku, Tokyo, 154-8525, Japan
Phone: +81-3-3418-9486
E-mail: paxton@komazawa-u.ac.jp

Tae Yamazaki, Associate Professor
Komazawa University
1-23-1 Komazawa, Setagaya-ku, Tokyo, 154-8525, Japan
Phone: +81-3-3418-9228
E-mail: ymzk@komazawa-u.ac.jp

Hannah Kunert, Assistant Professor
Komazawa University
1-23-1 Komazawa, Setagaya-ku, Tokyo, 154-8525, Japan
Phone: +81-3-3418-9582
E-mail: hannah80@komazawa-u.ac.jp

Received: September 30, 2022

Revised: November 23, 2022

Accepted: November 30, 2022