

**Investigating the consequential validity of TEAP: Washback to high school learners of English**

英語能力試験が受験者の勉強へ与える影響に関する調査研究

**David Allen**, Associate Professor, Ochanomizu University, Japan

**Diane Hawley Nagatomo**, Professor, Ochanomizu University, Japan

This report should be cited as follows:

Allen, D. & Nagatomo, D.H. (2019). *Investigating the consequential validity of TEAP: Washback to high school learners of English*. Tokyo: Eiken Foundation of Japan.

## Table of contents

Summary.....	4
1. Introduction .....	5
1.1 Washback .....	7
1.2 Test preparation strategies.....	10
1.3 Mediating factors of washback .....	10
1.4 The present study .....	12
2. Method .....	13
2.1 Participants.....	13
2.2 Procedure and materials.....	13
3. Results.....	17
3.1 Score data .....	17
3.2 Survey data.....	18
3.2.1 General English study.....	19
3.2.2 Preparation sessions .....	19
3.2.3 Motivation .....	20
3.2.4 Perceived difficulty .....	20
3.2.5 Expectancy of success .....	21
3.2.6 Perceived test value.....	21
3.2.7 TEAP preparation: Skills .....	21
3.2.8 TEAP preparation: Materials .....	22
3.2.9 TEAP preparation: Social and analytic strategies .....	22
3.2.10 TEAP preparation: Sub-skills and activities .....	23
3.2.11 Views on speaking tests .....	23
3.3 Interview data .....	26
3.3.1 Interviewees test data .....	26
3.3.2 The first interview.....	26
3.3.3 The second interview .....	28
3.3.4 The final interview .....	32
3.3.4 Teacher Interview .....	36
4. Discussion.....	40
4.1 Research question 1.....	40
4.2 Research questions 2 and 3 .....	41
4.3 Research question 4 .....	49

5.	Conclusions .....	51
6.	Recommendations.....	53
7.	References.....	54
8.	Appendices.....	59

## Summary

The present study investigates the consequential validity of the TEAP test from the perspective of washback on learning. Forty-six high school English learners took the TEAP test and completed a pre- and post-test survey examining their language learning and test preparation behaviors and perceptions regarding TEAP. Five of these participants were interviewed three times to attain further insight into language learning perceptions and behaviors. The learners' primary English language teacher was also interviewed in order to consolidate our interpretations of the data. The data show that participants were typically at the B1 level on the CEFR scale in all skills, both productive and receptive. This notable outcome was the result of their high school classes which focused greatly on developing learners' communicative competence, providing a balanced approach to teaching the four skills. Positive washback from the test was identified in terms of specific sub-skills (synthesizing ideas in writing, writing about visual data, reading and writing under time-constraints). Moreover, learner perceptions of the test were positive with learners drawing a distinction between the skills tested on the TEAP test and those tested on traditional entrance exams. The learners were also in favor of having speaking tests generally as part of entrance exams. Other findings revealed a lack of focus on note-taking and participating in role-plays in preparation for the test, even though these skills are believed to be tested in TEAP. Additionally, while we did observe positive washback, it was limited by the perceived importance of the test; most learners did not require TEAP scores for entrance purposes, which reduced the amount of preparation they did for the test. Also, positive washback on productive skills may have been reduced due to the focus on these skills in schools. We consider that the micro-context of the school in this study is perhaps not representative of that of the majority of high schools in Japan. In the broader context of Japanese high schools, we believe TEAP has the potential to achieve greater overall positive washback on productive skills, particularly in schools where such skills are not the focus of English classes. Moreover, washback is likely to be greatest for those learners that require TEAP scores for university entrance purposes.

## 1. Introduction

The Test of English for Academic Purposes (TEAP) was designed and developed collaboratively by the Eiken Foundation of Japan and Sophia University in Japan, and with consultants at the Centre for Research in English Language Learning and Assessment (CRELLA) at the University of Bedfordshire in the U.K. The TEAP Test is a four skills balanced test of academic English proficiency aimed at Japanese high school leavers and for use in university admissions and selection. The test was introduced in 2014 and since then it has been adopted by numerous universities for admissions purposes.

As part of its development a series of validation studies have been conducted by researchers at CRELLA focusing on the reading and listening (Taylor, 2014), speaking (Nakatsuhara, 2014; Nakatsuhara, Joyce & Fouts, 2014) and writing sections (Weir, 2014). The studies utilize the socio-cognitive model (Weir, 2005; O'Sullivan & Weir, 2011), which has been used particularly in the validation process of Cambridge examinations (Weir, Vidaković & Galaczi, 2013), as a framework for investigating the validity of the test. The model includes components relating to context, cognitive (theory-based), scoring, consequential and criterion-related validity, and to offer an acceptable validity argument to stakeholders, evidence must be gathered in support of each of the validity components (Weir, 2005). The validation studies by Taylor, Weir, and Nakatsuhara and colleagues provided initial evidence that the TEAP test demonstrated acceptable context, cognitive, and scoring validity.

In addition to these components of validity, evidence of consequential and criterion-related validity must also be sought and evaluated once the test is implemented in the context for which it was intended. Criterion-related validity includes concurrent and predictive validity and is primarily concerned with the extent to which test scores correlate with an appropriate external measure of performance (Messick, 1996). A recent study by In'nami, Koizumi and Nakamura (2016) presented an item-level confirmatory factor analysis with participants' scores on the TEAP and TOEFL (iBT) tests. The study found strong alignment of the

theoretical constructs underlying the two tests, thereby providing positive evidence for the concurrent validity of the TEAP test.

Consequential validity concerns the *consequences* of introducing a test, in other words, the *impact* that a test has on society. *Washback* fits under the umbrella of test impact but is much more specific in focus and concerns the effect that the test has upon teaching and learning. Regarding consequential validity, a large-scale study that is made up of four separate projects has been proposed to investigate the impact of the TEAP Test (Green, 2014). The aim of the first project was to produce an impact statement summarizing the types of impact that the test designers wished to achieve through the introduction of the test. This impact statement describes the intended social impact on the university entrance exam system and on values related to English, and the intended washback on EFL education in terms of both school-based teaching and learning and independent learning. The second project investigated the views of high school teachers and students on the state of current English education and the perceived consequences of introducing TEAP into the system. The third project, which is yet to be initiated, aims to investigate washback in context and to evaluate the impact on the teaching and learning practices following introduction of the test. The final project will seek to evaluate the outcomes of preparing for the test in terms of whether and to what extent learners' abilities actually improve as a result of the introduction of the test.

The present study bears resemblance to project three of the TEAP impact study described above. In this study, we investigate the consequential validity of the TEAP Test, specifically focusing on washback to the learner. In particular, we are looking at washback in terms of autonomous learning in the high-school context. The ultimate aim is therefore to investigate the extent to which positive washback is engineered in the Japanese high school context. This research thus complements the large-scale TEAP impact study by providing rich qualitative and quantitative data from a small-scale, context-dependent evaluation of washback from the TEAP Test. Prior to outlining the design of the present study, we provide a brief overview of relevant research theory and findings that will help us to contextualize and interpret our findings.

### 1.1 Washback

To demonstrate the consequential validity of a test, evidence must be provided that answers the question, 'What effects does the test have on its various stakeholders?' (O'Sullivan & Weir, 2011). As part of this, washback, which is the effect of a test upon teaching and learning (and teachers and learners), must also be evaluated on the basis of evidence. Both the direction of washback (positive or negative) and its intensity (Cheng, 2005) should be determined.

Washback begins with the design of the test (Saville, 2010) and is essentially derived from successful realization of construct validity (O'Sullivan and Weir, 2011, pp.21-22). Thus, whether a test is likely to create positive or negative washback is determined initially by the overlap of test items with the Target Language Use (TLU) domain (Green, 2007; Messick, 1996). As previously noted, research has demonstrated acceptable construct validity for the TEAP Test (Nakatsuhara, 2014; Nakatsuhara et al., 2014; Taylor, 2014; Weir, 2014), which constitutes the first important step in attempting to generate positive washback. However, washback is manifested through the *interaction* between the test and the various participants, processes and products (Bailey, 1996; Hughes, 2003), and thus researchers must provide evidence of washback in contexts in which the tests are actually being used.

It has been observed from the outset that washback is a highly complex process that occurs in real-life contexts involving numerous stakeholders, processes and outcomes (Alderson & Wall, 1993; Bailey, 1996; Hughes, 2003; Watanabe, 2004). Of the numerous stakeholders, the learners are arguably those whose lives are most affected by the test. It is somewhat surprising therefore that the majority of early washback studies tended to focus primarily on the effects of tests upon teachers and teaching rather than learners and learning (Cheng, 2014; Watanabe, 2004). More recently, however, the number of studies focusing on washback to learners and upon learning has increased considerably (e.g., Allen, 2016a, 2016b, 2017; Cheng, Andrews & Ying, 2011; Gosa, 2004; Green, 2005, 2006, 2007; Mickan & Motteram, 2009; Pan, 2014; Shih, 2007; Smyth & Banks, 2012; Stoneman, 2006; Tsagari, 2007; Tsai & Tsou, 2009; Xie, 2013; Xie & Andrews, 2012; Zhan & Andrews, 2014; Zhan & Wan, 2016). The findings of these studies have led to a much more sophisticated view of how washback affects the learner and learning.

The study reported in Allen (2016a, 2017) is of particular importance for the present research because of its context, orientation and design. The study investigated the washback effect from the newly-introduced IELTS Academic Test in the Japanese tertiary context. Although the IELTS Test is one of the most established English for Academic Purposes (EAP) tests, the majority of the participants in the study were completely unfamiliar with the test and with the tasks that it is composed of. This situation allowed the researcher to investigate the impact of the test upon the learner and upon learning. In the study, 190 undergraduates completed two IELTS Tests over a two-year period, the second of which was followed by a comprehensive survey that detailed participants' language learning experience prior to university (i.e., at high school and cram school, which is known as *juku* or *yobiko* in Japan), at university, and most importantly in preparation for the two IELTS Tests. In addition, 19 participants were interviewed to gain a richer understanding of their learning experience and test preparation behaviour.

A number of findings from that study are particularly relevant to the present one. Firstly, regarding the test scores, it was shown that the undergraduates were much more proficient in reading (Test 1 Mean score=7.2) and listening (M=6.4) than writing (5.5) and speaking (5.4). Through a synthesis of the survey and interview data, as well as a general overview of the contemporary English educational context in Japan, it became clear that these results reflected a bias towards learning receptive skills at high school (also see Green, 2014), and particularly at cram school (also see Allen, 2016b). In other words, a strong washback effect of the high-stakes university entrance exam was observed. This finding confirmed long-standing intuitions about the likely influence of Japanese university entrance exams upon pre-tertiary learning practices and learners' resulting language abilities (e.g., Brown & Yamashita, 1995; Kikuchi, 2006).

Another important finding was that the IELTS Test appeared to stimulate learning of the productive skills in this context. This was evidenced through multiple data sources. Firstly, learners' speaking test scores significantly increased over the period. Secondly, survey data revealed that learners focused significantly more on productive skills during test preparation, while reducing their focus on reading. Thirdly, the interview data revealed that learners perceived the productive



component tasks to be the most difficult and unfamiliar, and this increased awareness was a driving force behind their focus on productive skills. These findings were in line with the observed increase in speaking proficiency and suggested a positive washback effect on productive skills from the IELTS Test upon learning in the Japanese context.

However, there were numerous mediating factors that influenced the strength and direction of the washback effect. In particular, because learners had had little experience of writing academic texts as required in the IELTS Test and also because they had little opportunity to develop the speaking skills necessary for participating in the oral interview, they were often at a loss for how to prepare for such tasks. Although the test preparation materials used played an important role in facilitating positive washback (also see Saville & Hawkey, 2004), learners reported difficulty and dejection when attempting to prepare for the productive tasks and study productive skills. This was partly due to the fact that learners prepared for the test in a non-instructed context (i.e., not attending test preparation courses) but was also due to their previous learning experiences. The research thus revealed a number of mediating factors that shaped washback to the learner.

The findings reported above are particularly important for the present study for a number of reasons. Firstly, both studies share a broad context (i.e., Japanese pre-tertiary and tertiary English education). Secondly, both studies involve learners who are preparing for a test in a non-instructed context. Thirdly, both the IELTS and TEAP Tests are four skills tests of academic English language ability, though they are distinct in many ways, notably in terms of the test tasks utilized and the intended use (i.e., IELTS is intended to be used for immigration and admission to English-medium universities around the world, while the TEAP Test is designed specifically for use in the admissions process of Japanese universities). Finally, both studies focus on washback on learning, rather than teaching.

In the following section, we define the focus of this study and review a number of the key findings that are likely to be important to understand the washback process from the TEAP Test in the Japanese context.

### *1.2 Test preparation strategies*

One primary focus of this study is how the TEAP Test influences test takers' preparation strategies, that is, the things that they do in preparation for the test. If the test leads learners to prepare in ways that are likely to be positive for their learning of the target language, this can be taken as evidence of positive washback upon test preparation practices. To determine whether learning behaviour is directly influenced by the test, evidence that shows a *change* in this behaviour must be provided (Messick, 1996). To do this, we must investigate the way that learners currently study English and the way that they prepare for the test; by comparing the two, we can potentially attribute changes to the introduction of the test.

Research investigating test preparation in instructed contexts (i.e., when learners are attending test preparation classes) shows that students tend to focus on test-related tasks and materials when preparing for a test. In other words, they narrow their focus of study to the content that they perceive to be relevant to the test (Gosa, 2004; Green, 2007; Mickan & Motteram, 2009; Shih, 2007; Stoneman, 2006; Wall & Horak, 2011; Xie, 2013; Zhan and Andrews, 2014; Zhan & Wan, 2016). Therefore, although test materials can have a considerable impact on what and how learners prepare for a test (Saville & Hawkey, 2004), the variability inherent in learners' perceptions of the test means that it cannot be taken for granted that all learners will use the materials in the same way and to the same effect. This is especially the case when learners prepare for a test in non-instructed contexts. Consequently, it is necessary to investigate which skills learners practice in preparation for the test, rather than taking it for granted that learners will practice those skills intended by the test and materials designers.

### *1.3 Mediating factors of washback*

It is also important to investigate the reasons behind learners' choice of test preparation behavior. By doing so, it is possible to evaluate the factors that may mediate washback to the learner. Previous studies have shown that beliefs, educational experience and contextual circumstances can all mediate washback to learners and learning (Allen, 2016a, 2017; Gosa, 2004; Green, 2007; Xie & Andrews, 2012; Zhan & Andrews, 2014; Zhan & Wan, 2016). Perceptions of test difficulty as

well as the perceived importance of the test also affect the washback generated: for optimal washback, a test should be both important and challenging (Green, 2007).

One of the most salient learner-related issues in the present context is that of preparing for the productive skill components. It was noted earlier that in the Japanese English education system, receptive skills tend to dominate both teaching and learning activities (Allen, 2017; Green, 2014). Consequently, when faced with a balanced four skills test, many Japanese learners may have difficulty in adjusting themselves to the demands of the test. This is related to both their knowledge and ability about how to practice and improve their language production, as well as the resources available to them, in the form of materials, peers and teachers. Previous research has shown that learners often lack the knowledge about how to study for speaking and writing and this is perhaps due to a lack of focus on these skills in the local educational context (Allen, 2016a, 2017; Shih, 2007). Learners may also be overly-dependent on others for assistance and feedback, especially regarding the productive skills (Allen, 2016a, 2017; Mickan & Motteram, 2009). Some learners may thus lack agency and self-direction in the face of assessment (Mickan & Motteram, 2009). Importantly, learners who are preparing for a test independently are presumably more likely to experience these difficulties than those who are participating in test preparation courses where teachers (and peers) can help them develop autonomy as well as provide feedback on their language production.

In sum, it is necessary to investigate not only what learners do in preparation for a test (i.e., test preparation strategies) but also why they do it (i.e., the factors that influence the adoption or rejection of particular strategies). Because the explicit aim of the TEAP Test is to stimulate positive washback on English education in Japan, both in the classroom and during autonomous learning, it is essential to investigate the test preparation strategies that students adopt and the factors that mediate their choices of strategies. In this way, it is possible to evaluate the consequential validity of the TEAP Test, specifically whether and to what extent it leads to positive washback on learners' test preparation.

#### *1.4 The present study*

In this study, we investigated *how* high school students from one high school in Japan prepared for the TEAP test and *why* they prepared in that way. We conducted pre- and post-test surveys to investigate how the students study English in general (i.e., before the test was introduced) and specifically for the test. In addition, we conducted interviews with a subset of the participants to investigate in more detail learners' approaches to studying and test preparation and views of the test. The following research questions were formulated:

1. What were the scores for the TEAP Test and for each component skill? Was there a bias towards receptive skills as seen in other studies with other tests in the Japanese context?
2. Do test takers' preparation strategies differ from those that they normally employ when learning English? If so, does the change indicate positive washback on learning processes?
3. What factors mediate the process of washback from the TEAP Test?
4. What were test takers' views on the TEAP Test and their experience of preparing for it? Did they feel the test helped them to improve their English?

## 2. Method

### 2.1 Participants

Initially, sixty-four second-year students (aged between 16 and 17) at an all-girls Japanese high school in the Tokyo metropolitan area were recruited for the study. All participants completed the first survey. Fifty of these participants were then selected based on whether they had intended to take the test for their own goals (i.e., for university entrance purposes) and whether they were likely to be able to take the test and complete the second survey. Of these, five participants were also recruited for interviews, with preference given to those who were likely to need the TEAP scores in the future. Forty-six of the participants completed the test and both surveys (i.e., 92% completion rate).

All participants gave their consent to participate in the study and the study was approved by the research ethics committee at Ochanomizu University. The TEAP test fees were covered by the Eiken Foundation of Japan, enabling all participants to take the test, while interviewees were paid a small additional sum for participation. One student reported having taken TEAP previously, while the others had no prior experience of the test.

The participants in this study were all high academic achievers, having succeeded in gaining entry to a prestigious national high school. Consequently, students entering this high school may understand ‘the rules of the game’ (Bourdieu, 1990); in other words, they had a good understanding of what is required of them in exam-oriented educational systems, an attribute often associated with aspiring middle-class students (e.g., Smyth & Banks, 2012).

### 2.2 Procedure and materials

The procedure for the research involved five stages: First, participants completed the initial online survey in October 2016. The survey instruments were created and administered online using *Survey Monkey* and were written in the participants’ first language, Japanese. The choice and construction of the items was informed by previous studies investigating washback on test preparation practices, particularly Allen (2017), Green (2007) and Xie and Andrews (2012). The pre-test survey contained seven sections with 93 items (four open, five categorical and 84 Likert

scale response items; see Appendix 1 for survey items). Likert items were all on a 6-point scale of agreement (1=*strongly disagree*, 2=*disagree*, 3=*somewhat disagree*, 4=*somewhat agree*, 5=*agree*, 6=*strongly agree*). Table 1 summarizes the content of the pre- and post-test surveys.

*Table 1: Summary of survey content*

	Pre-test (survey 1)	Post-test (survey 2)
<ul style="list-style-type: none"> <li>• Previous and current English language learning: <ul style="list-style-type: none"> <li>- Age began learning, study abroad experience</li> <li>- Perceived proficiency</li> <li>- Intensity of current language study, use of English outside of class</li> <li>- Skill focus at school</li> <li>- Skill focus during independent study</li> <li>- Activities and sub-skills related to studying English</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>✓</li> <li>✓</li> <li>✓</li> <li>✓</li> <li>✓</li> <li>✓</li> </ul>	<ul style="list-style-type: none"> <li>✓</li> </ul>
<ul style="list-style-type: none"> <li>• Perception of test importance (value) <ul style="list-style-type: none"> <li>- Perception of the value of English for their future</li> <li>- Perception of the value of TEAP for their future</li> <li>- Views on speaking test</li> <li>- Expected/Perceived improvement as a result of preparing for TEAP</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>✓</li> <li>✓</li> <li>✓</li> <li>✓</li> </ul>	<ul style="list-style-type: none"> <li>✓</li> <li>✓</li> <li>✓</li> <li>✓</li> </ul>
<ul style="list-style-type: none"> <li>• Preparation sessions <ul style="list-style-type: none"> <li>- Impact of the preparation sessions on preparation behavior for TEAP</li> </ul> </li> </ul>		<ul style="list-style-type: none"> <li>✓</li> </ul>
<ul style="list-style-type: none"> <li>• Motivation</li> </ul>	<ul style="list-style-type: none"> <li>✓</li> </ul>	<ul style="list-style-type: none"> <li>✓</li> </ul>
<ul style="list-style-type: none"> <li>• Perception of test difficulty</li> </ul>	<ul style="list-style-type: none"> <li>✓</li> </ul>	<ul style="list-style-type: none"> <li>✓</li> </ul>
<ul style="list-style-type: none"> <li>• Expectation of success</li> </ul>	<ul style="list-style-type: none"> <li>✓</li> </ul>	
<ul style="list-style-type: none"> <li>• Test preparation <ul style="list-style-type: none"> <li>- Number of hours studied</li> <li>- Practicing test related (four) skills</li> <li>- Materials used</li> <li>- Socio-affective strategies (e.g., asking others for advice)</li> <li>- Test taking strategies (e.g., analyzing test papers)</li> <li>- Activities and sub-skills related test preparation</li> </ul> </li> </ul>		<ul style="list-style-type: none"> <li>✓</li> <li>✓</li> <li>✓</li> <li>✓</li> <li>✓</li> <li>✓</li> </ul>

In addition, five participants took part in semi-structured interviews. The interviews were conducted in Japanese by a trained postgraduate research assistant. We felt that students would speak more freely to her about their preparation strategies. Interview questions were similar to those in the survey, but interviewees were encouraged to discuss the reasons for their test preparation strategies (see Appendix 2 for the questions used to guide the semi-structured interviews). In addition, the interviewer referred to the participants' survey responses as necessary.

Next, all participants took part in two three-hour preparation sessions in November. These sessions were conducted by the researchers at Ochanomizu University and were provided to all second-year students at the high school (including those not participating in the study). The sessions focused on all parts of the TEAP Test, particularly the productive skill components, as these were assumed to be more challenging for the students. A number of questions were included in the second survey to investigate the impact of these sessions on students' test preparation behavior.

After the sessions, participants were instructed to prepare individually for the test. The researchers had no further contact with participants prior to the test.

Approximately six months later, in July 2017, participants took the TEAP test at the officially designated test center in the Tokyo. The test takers received their results individually in the standard format for test takers and as the participants agreed in their consent forms, we were provided with the score data courtesy of the Eiken Foundation of Japan.

Following the test, participants completed the second online survey. The majority of participants completed the second survey within two weeks and the remainder completed the survey within four weeks of completing the test. The post-test survey contained six sections with 105 items (seven categorical, nine open and 89 Likert scale response items). Seventy-three items were repeated for the first and second test to facilitate pre- and post- test comparisons.

In addition to the final survey, the five interviewees were interviewed individually a second time . These were held within a week of completing the test.

A third round of interviews with the same five students was conducted again in March 2018, just prior to entering university, to further establish their views on the TEAP test.



### 3. Results

#### 3.1 Score data

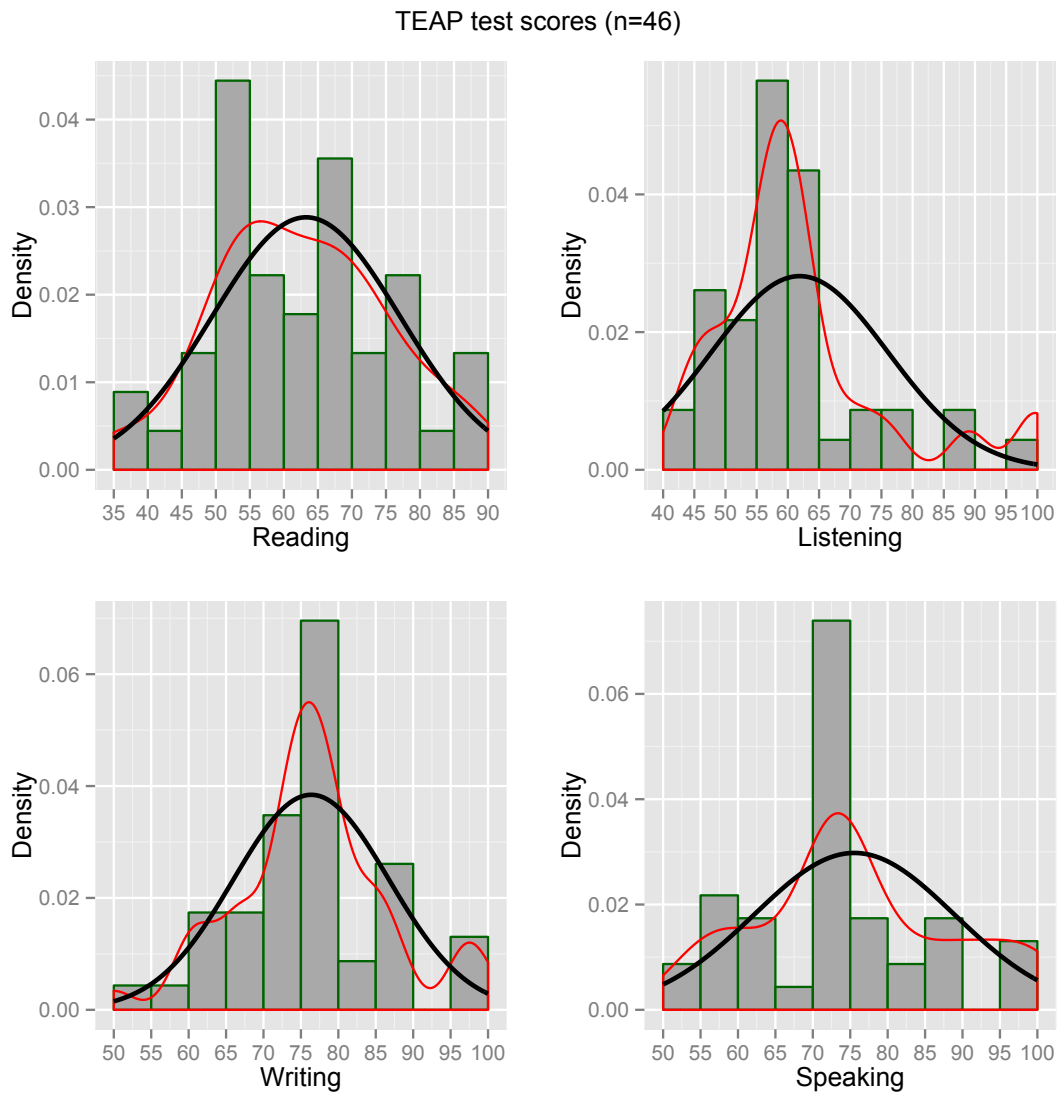
The TEAP score data are summarized in Table 2 and Figure 1. Mean scores were higher for productive skills and the score ranges for these skills were more restricted than those for receptive skills. In terms of the CEFR proficiency scale, mean scores for each skill were at the B1 level falling roughly in the middle of the score range for that band (i.e., Listening: 51~74; Reading: 50~74; Writing and Speaking: 61~84).<sup>1</sup>

*Table 2: Descriptive statistics of TEAP score data*

	CEFR	M	SD	Min	Max	Skew	Kurtosis	SE
Reading	B1	63.3	13.8	35	100	0.3	-0.1	2.0
Listening	B1	61.9	14.2	43	100	1.3	1.2	2.1
Writing	B1	76.4	10.4	50	100	0.1	0.2	1.5
Speaking	B1	75.5	13.4	52	100	0.2	-0.8	2.0
Overall	B1	277	45.4	202	376			

---

<sup>1</sup> The CEFR benchmarked score ranges are available from the Eiken Foundation of Japan website [https://www.eiken.or.jp/teap/info/2018/0418\\_01.html](https://www.eiken.or.jp/teap/info/2018/0418_01.html). Note these are the ranges for 2017 when the test was taken; the score ranges since have been updated for tests taken from 2018.



*Figure 1: TEAP test scores (n=46) for the four skills. Black lines denote the normal distribution curve and red lines show how the present data deviates from this.*

### 3.2 Survey data

The results of the survey are organized by topic. Where possible, Likert scale responses to the survey questions were compared for the two tests. The data were not normally distributed, however, and therefore Wilcoxon Signed-rank tests were performed with  $r$  as the effect size (.1=small, .3=medium, .5=large). A Bonferroni correction for multiple tests was applied to allow for a more conservative estimation of statistical significance ( $p$  at  $.05/73 = p < .007 = *$ ;  $.01/73 = p < .0001 = **$ ). Statistically significant results are noted in the text. Cronbach's estimate of internal consistency for 167 Likert scale items in the surveys was high ( $\alpha = .95$ ; with the items in the first and second surveys, reaching  $\alpha = .94$  and  $\alpha = .93$ , respectively).

### *3.2.1 General English study*

In the first survey, participants were asked about their general English learning. They reported beginning to learn English when they were between 12 and 15 years old ( $n=30$ ), between seven and 11 years old ( $n=19$ ), between three and six ( $n=11$ ) and from birth to two years old ( $n=4$ ). Ten participants reported having spent time abroad, with duration of stay ranging from one to 10 months. Participants' self-rated proficiency on a ten-point scale (0-9) was 4.3 ( $SD=1.4$ ) on the first survey and this increased slightly on the second survey ( $M=5.0$ ,  $SD=1.4$ ).

There was strong agreement that English was useful for their future ( $M=5.4$ ,  $SD=0.9$ ). Of the four skills, speaking was thought to be the most important ( $M=5.6$ ,  $SD=0.6$ ), closely followed by listening ( $M=5.4$ ,  $SD=0.8$ ), reading ( $M=5.3$ ,  $SD=0.9$ ) and writing ( $M=4.9$ ,  $SD=0.9$ ).

Most participants reported studying English for either four or six hours per week ( $n=14$ ,  $n=11$ , respectively), with a handful studying less (2 hours,  $n=7$ ; <1 hour,  $n=3$ ) or more than this (8 hours,  $n=5$ ; 10 hours,  $n=4$ ; 10+hours,  $n=2$ ). Responding to the question 'how often do you use English outside of class', large proportions varied between 'not at all' ( $n=22$ ), 'not very much' ( $n=19$ ), 'often' ( $n=18$ ), with much fewer responding 'almost every day' ( $n=5$ ) and 'every day' ( $n=1$ ).

Participants reported spending most of their time in English classes learning speaking ( $M=5.0$ ,  $SD=1.0$ ), followed by listening ( $M=4.3$ ,  $SD=1.0$ ), reading ( $M=3.8$ ,  $SD=1.0$ ), and finally writing ( $M=3.6$ ,  $SD=1.1$ ). In contrast, participants spent most of their independent English study time reading ( $M=4.5$ ,  $SD=1.2$ ), followed by writing ( $M=3.7$ ,  $SD=1.3$ ) and listening ( $M=3.3$ ,  $SD=1.2$ ), and finally speaking ( $M=2.4$ ,  $SD=1.1$ ).

### *3.2.2 Preparation sessions*

Two Likert-scale questions targeted participants' perceptions of the impact of the preparation sessions conducted by the researchers. Overall, participants thought the sessions helped them to understand the TEAP test ( $M=4.0$ ,  $SD=1.3$ ). Participants were most motivated to study writing after the sessions ( $M=4.1$ ,  $SD=1.3$ ), followed by speaking ( $M=3.8$ ,  $SD=1.3$ ), reading ( $M=3.7$ ,  $SD=1.2$ ) and listening ( $M=3.7$ ,  $SD=1.3$ ). However, participants generally disagreed that they had changed their approach to studying the four skills ( $M=2.2\sim2.6$ ,  $SD=1.1\sim1.4$ ).

Open-ended written responses, on the other hand, illustrated various ways in which participants changed their approaches to studying the four skills following the preparation sessions and while studying for TEAP (Appendix 3). Twenty-two participants provided 53 specific textual responses and these were distributed fairly evenly across the four skills (S=12, W=14, R=14, L=13).

Regarding speaking, participants reported speaking within the allotted time, thinking of paraphrases and synonyms, thinking of reasons and opinions, role-playing the interview (part 2), speaking more with others, using past papers, recording responses, reading aloud and shadowing from a CD.

For writing, participants reported increasing the amount and frequency of study, studying composition and text structure, checking university-related vocabulary, practicing summarizing information rather than structuring one's own opinions, focusing on the coherence and objectivity of argumentation, reviewing and editing, and revising grammar.

For reading, participants reported studying academic vocabulary, reading more, reading longer texts, paying attention to specific reading skills, reading quickly to identify the gist, trying to grasp the main idea of paragraphs and attending to information needed when reading (i.e., the purpose of reading).

Participants reported listening more frequently to English, the radio, and various genres of English, using TEAP-specific listening materials and/or past papers, note-taking, and reading the script while re-listening.

### *3.2.3 Motivation*

Participants were highly motivated to study the four skills prior to the test ( $M=4.6\sim4.7$ ,  $SD=0.8\sim1.1$ ) and TEAP motivated them to study the four skills more in the future ( $M\sim4.5\sim4.9$ ,  $SD=1.0\sim1.1$ ).

### *3.2.4 Perceived difficulty*

Participants tended to agree that the TEAP test would be difficult, with speaking ( $M=4.8$ ,  $SD=1.2$ ) and listening ( $M=4.7$ ,  $SD=1.1$ ) predicted to be slightly more difficult than reading ( $M=4.4$ ,  $SD=1.0$ ) and writing ( $M=4.3$ ,  $SD=0.9$ ). After the test, participants thought that speaking was the most difficult ( $M=4.7$ ,  $SD=1.3$ ), followed

by listening (M=4.4, SD=1.1), reading (M=4.2, SD=1.1) and finally writing (M=3.6, SD=1.0). Notably, writing was perceived to be significantly less difficult than predicted ( $p<.006^*$ ,  $r=.38$ ).

### *3.2.5 Expectancy of success*

Participants generally felt that they would be able to succeed on the TEAP test; however, they felt less confident about speaking (M=3.7, SD=1.2) than on other parts of the test (reading M=4.7, SD=1.1; writing M=4.6, SD=1.0; listening M=4.4, SD=1.0).

### *3.2.6 Perceived test value*

Ten participants thought TEAP would be useful for their future, 30 were unsure and six thought it would not. Following the test, more participants felt that the test would *not* be of use to their future ( $n=16$ ), while 22 were still unsure and eight thought it would. This change suggests students had moved closer towards deciding which universities they would be applying to, many of which may not accept TEAP scores.

Participants were asked whether they thought that preparing for the test would help them to improve their English (Survey 1) and whether preparing for the test actually did so (Survey 2). The comparison showed that participants were considerably more hopeful of improvement than their perceived improvement subsequent to the test (M=4.8, SD=0.9; M=3.4, SD=1.3;  $p<.0001^{**}$ ,  $r=.54$ ).

### *3.2.7 TEAP preparation: Skills*

Regarding the number of hours that they reported studying for the test, 32 participants prepared for less than 20 hours, eight studied between 20 and 40 hours, one studied 40-60 and another 60-80 hours, and four reported not studying at all.

Participants reported preparing most for reading (M=3.6, SD=1.4), writing (M=3.5, SD=1.3), listening (M=3.3, SD=1.4) and least for speaking (M=2.9, SD=1.4; Figure 2). The order of priority matches that of participants' independent study. Participants reported studying speaking more for the test than they normally do in their independent study (M=2.9 vs. M=2.4), though this difference was not

significant, and speaking was the least practiced of the four skills. Although reading was the most studied skill in both independent study and test preparation, participants reported studying reading significantly less for the test ( $M=4.5$  vs.  $M=3.6$ , respectively;  $z=2.997$ ,  $p<.006^*$ ,  $r=.31$ ).

Participants reported studying speaking and listening significantly more in class than in preparation for the test ( $M=5.0$  vs.  $M=2.9$ ,  $z=5.524$ ,  $p<.0001^{**}$ ,  $r=.58$ ;  $M=4.3$  vs.  $M=3.3$ ,  $z=3.584$ ,  $p<.006^*$ ,  $r=.37$ ), while there was no such difference for reading and writing.

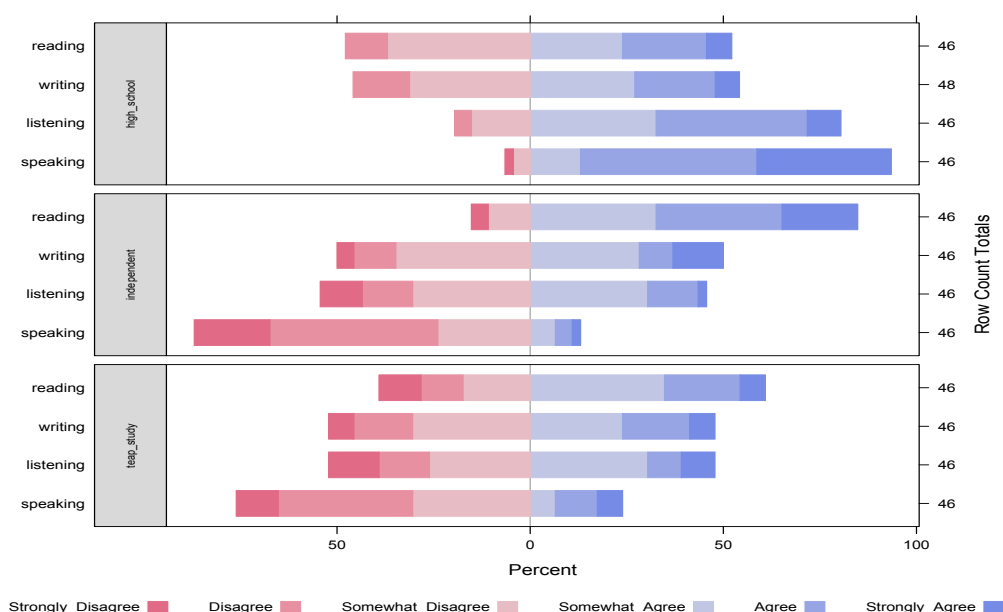


Figure 2: Focus on skills at high school, during independent study and for TEAP

### 3.2.8 TEAP preparation: Materials

Most participants used TEAP reference books to prepare ( $n=32$ ) while a few used TEAP-related online resources ( $n=4$ ). Three participants reported using NCUEE exam materials and one used EIKEN Level Pre-1 test materials. The average number of full mock TEAP tests done in preparation was 0.8 ( $SD=0.8$ ). Most participants reported doing one test ( $n=22$ ) or none ( $n=18$ ), followed by two ( $n=5$ ) and three tests ( $n=2$ ).

### 3.2.9 TEAP preparation: Social and analytic strategies

Participants tended not to seek advice about English or the test ( $M=2.7$ ,  $SD=1.6$ ) and did not seek feedback from others about their English speaking ( $SM=2.5$ ,  $SD=1.7$ ) or

writing ( $M=2.7$ ,  $SD=1.5$ ). They also tended not work together to give and receive feedback on writing ( $M=2.3$ ,  $SD=1.3$ ) or speak with others in English ( $M=2.7$ ,  $SD=1.5$ ). Participants reported midline agreement about whether they analyzed test questions ( $M=2.9$ ,  $SD=1.4$ ) and studied the frequency of the different question types ( $M=2.9$ ,  $SD=1.4$ ), indicating that this was not a major focus of test preparation.

### *3.2.10 TEAP preparation: Sub-skills and activities*

Responses to items that focused on the knowledge and skills that students focus on during their general English study (Survey 1) and in preparation for TEAP (Survey 2) are shown in Table 3 (Additional figures in Appendix 4). These comparisons were significant for studying both vocabulary and grammar, demonstrating that participants studied these aspects much more in general than for the test ( $M=3.8\sim4.0$  vs.  $M=2.7\sim3.1$ ). Of the four skills, participants appeared to exhibit the greatest variation and change in behaviors for reading and writing. Reading and writing under time pressure were done significantly more for the test ( $M=3.1$  vs.  $M=3.8$ ;  $M=3.0$  vs.  $M=4.0$ ), while reading stories was done significantly less ( $M=4.3$  vs.  $M=2.5$ ). Writing summaries of texts and visual materials (i.e., synthesizing various sources) and writing about information in graphs and charts were done significantly more for TEAP ( $M=2.2$  vs.  $M=3.2$ ;  $M=2.2$  vs.  $M=3.1$ ). Participants practiced note-taking and oral role-plays significantly less for the test ( $M=3.1$  vs.  $M=2.0$ ;  $M=3.7$  vs.  $M=2.7$ ).

### *3.2.11 Views on speaking tests*

Participants tended to agree that speaking tests were needed on English entrance exams ( $M=4.2$ ,  $SD=1.5$ ; Figure 3). Most participants ( $n=38$ ) commented on this question (*Should university entrance exams test English speaking?*) (Appendix 5). Five participants expressed disagreement citing reasons of the perceived difficulty of assessing speaking in general, being able to assess speaking fairly, and the burden on students. The remaining 33 comments expressed positive views. The most prevalent reason was that participants wanted to be able to use English and that speaking was the most important skill for communication. Others stated that passing tests and being able to understand grammar were pointless if one cannot speak and that

being able to speak a language is the evidence of having acquired a language. On the whole, participants' comments expressed a strong desire to be able to communicate in English.

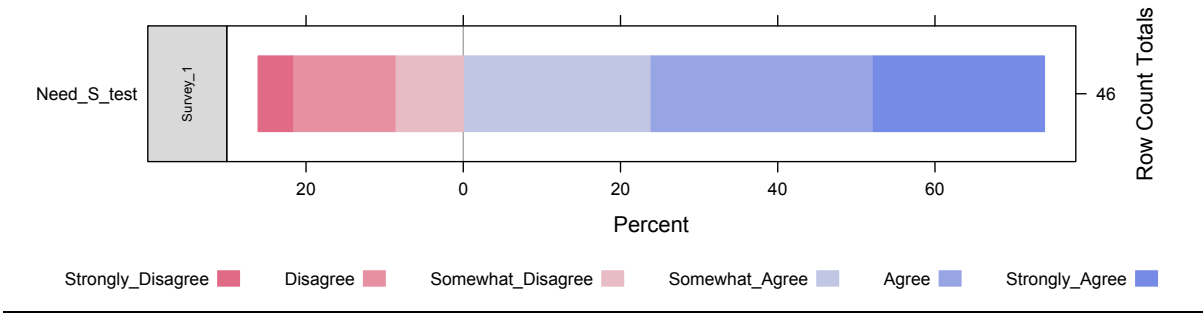


Figure 3: Responses to question 'Should university entrance exams test English speaking?'



*Table 3: Self-reported study of task-specific sub-skills*

	Survey 1		Survey 2		Wilcoxon Signed-rank test statistic, $p$ -value ( $p < .007 = *$ ; $p < .0001 = **$ ), and $r$ .
	M	SD	M	SD	
Academic vocabulary	3.4	1.3	3.2	1.3	
Vocabulary for reading and listening	4.0	1.0	3.1	1.4	$Z = 3.369, *, 0.35$
Vocabulary for speaking and writing	4.1	1.1	2.7	1.3	$Z = 4.386, **, 0.46$
Grammar for reading and listening	3.8	1.2	2.7	1.2	$Z = 4.239, **, 0.44$
Grammar for speaking and writing	3.9	1.1	2.7	1.2	$Z = 3.896, **, 0.41$
Careful reading	3.6	1.1	3.0	1.2	
Skim reading	4.1	1.2	4.3	1.1	
Search reading	4.0	1.1	4.4	1.0	
Reading various genres (blogs, emails, advertisements)	2.9	1.5	2.5	1.2	
Reading academic texts	2.8	1.2	3.0	1.2	
Reading educational materials	3.1	1.1	2.7	1.2	
Reading texts with visual information	3.0	1.1	3.2	1.3	
Reading stories	4.3	1.1	2.5	1.3	$Z = 5.488, **, 0.57$
Reading information about universities	2.0	1.2	2.0	1.0	
Reading fluently under time pressure	3.1	1.3	3.8	1.4	$Z = -3.341, *, 0.35$
Analysing answers to comprehension questions	4.4	1.1	4.7	1.1	
Reading everyday	3.4	1.4	3.9	1.2	
Note-taking	3.1	1.1	2.0	1.2	$Z = 4.392, **, 0.46$
Careful listening	4.1	1.0	4.0	1.3	
Gist listening	4.5	1.1	4.1	1.2	
Listening to lectures	3.3	1.4	2.8	1.6	
Listening to materials containing visual information	2.7	1.1	2.7	1.3	
Listening to two-person dialogues	3.5	1.3	3.4	1.4	
Listening to dialogues with 3 or more speakers	3.0	1.4	2.7	1.3	
Listening to audio on the internet	3.9	1.4	3.3	1.6	
Listening to television and movies	3.5	1.6	2.9	1.8	
Listening to information about universities	2.0	1.2	1.6	0.9	
Listening everyday	3.8	1.5	3.4	1.4	
Speaking about personal topics	3.6	1.2	3.0	1.5	
Speaking about abstract topics	2.6	1.0	2.7	1.5	
Answering questions	3.8	1.0	3.8	1.5	
Asking questions	3.4	1.2	3.2	1.4	
Giving my opinion	3.6	1.2	3.2	1.4	
Role-playing	3.7	1.4	2.7	1.6	$Z = 3.271, *, 0.34$
Speaking fluently	3.2	1.2	3.2	1.6	
Speaking and recording myself	3.1	1.4	2.6	1.4	
Trying to use new vocabulary	3.8	1.2	3.3	1.2	
Speaking accurately	3.5	1.1	3.3	1.3	
Using more complex grammar and vocabulary	2.3	0.9	2.3	1.2	
Pronunciation	4.1	1.2	3.8	1.5	
Intonation	4.1	1.1	3.6	1.6	
Speaking everyday	3.4	1.3	3.2	1.5	
Writing long texts	3.5	1.3	3.9	1.5	
Writing summaries	3.7	1.2	3.8	1.3	
Writing summaries of texts and visual materials	2.2	1.2	3.2	1.4	$Z = -3.433, *, 0.36$
Writing about visual materials	2.2	1.0	3.1	1.5	$Z = -3.060, *, 0.32$
Reviewing my writing	4.3	1.1	4.2	1.3	
Rewriting my work	3.7	1.1	3.6	1.4	
Using new vocabulary and in my writing	4.1	1.0	4.0	1.3	
Focusing on content and organization	4.1	0.9	4.3	1.1	
Writing fluently under time pressure	3.0	1.1	4.0	1.3	$Z = -3.622, *, 0.38$
Using complex grammar and vocabulary	2.8	1.1	3.1	1.2	
Writing accurately	4.0	1.0	4.0	1.3	
Writing everyday	3.0	1.2	3.0	1.3	

### 3.3 Interview data

#### 3.3.1 Interviewees test data

Table 4 shows the interviewees' individual scores (pseudonyms are used throughout). One interviewee (Kaori) did not take the test due to illness but we present her data nonetheless. The remaining four interviewees' overall scores were all higher than the average of the participants in the main study ( $M=277$ ). Kimie was the most proficient user (and the highest scorer in the whole sample) but scored lower on writing than Mari, whose score for writing was markedly higher than on the other sections. Scores for speaking were highest of the skills for three out of four of the participants.

Table 4: Interviewees raw TEAP scores and CEFR bands.

TEAP Scores									
	Reading		Listening		Writing		Speaking		Overall
Kimie	100	B2	100	B2	76	B1	100	B2	376
Noriko	81	B2	85	B2	73	B1	95	B2	334
Megumi	79	B2	75	B2	76	B1	95	B2	325
Mari	75	B2	70	B1	97	B2	73	B1	315

#### 3.3.2 The first interview

The first interviews were conducted in October 2016 immediately after completing the questionnaire and prior to attending the preparation sessions. They focused on the participants' current approach to English language study, their language learning backgrounds and goals.

Table 5 summarizes how the participants study at school, at cram school, and independently. The students' comments indicate a distinction between the high school focus of developing students' communicative proficiency in the four skills and the cram school's focus on *juken* [entrance examination] English. However, the students mentioned that the type of communicative activities they were doing in the high school as second year students were likely to be replaced with more *juken*-like (i.e., test preparation) activities after they entered the third grade. Independent study involved following teachers' recommendations, not only grammar and vocabulary workbooks, but also independent reading and writing projects. Participants also devised their own study methods. Kaori, for instance, instead of

memorizing grammar rules, practiced writing and then analyzed it for mistakes. Mariko purchased a grammar book and set a goal of many pages per week to do. All participants selected authentic listening materials that reflected their own interests.

*Table 5: Summary of how the interviewees' study at high school, cram school and independently*

Study focus	High school	Cram school	Independent study
Reading	Reading marathon Read foreign books	Read long texts aloud Study past entrance exams	Read English books Browse the internet Read aloud quickly Review what was learned in schools Follow Twitter, Instagram EIKEN study books
Writing	Opinion essays Free writing Diaries Summarize and write opinion on reading	Translating sentences and texts (in both directions) Summarizing long readings	Write down sentences, meanings, and collocations EIKEN study books
Listening	Listening activities related to text, TED Talks	Not mentioned	Foreign music Foreign movies TED Talks YouTube Podcasts EIKEN exercise books
Speaking	Shadowing Talk about topic written on blackboard Working with partner Pronunciation practice	Not mentioned	Attend conversation school Practice speaking alone Memorizing sentences and saying them out loud
Grammar	Focusing on meaning	Work through textbook Focusing on rules, grammatical terms, Translation	Self-study books (Risen) Preparing for G-TEC (school homework and extra self-study)
Vocabulary	Vocabulary book for homework Collocations	Vocabulary book Memorize translations Translate vocabulary in long texts Understand sentence structure	Vocabulary books (school homework and extra self-study) Look up unknown words when reading, listening EIKEN study books

All interviewees see English as a means to develop social and cultural capital; they would be able to travel widely, communicate with many kinds of people, and deepen their hobbies and interests (e.g., follow Cirque Du Soleil on Instagram and Twitter, watch movies without subtitles, and enjoy music). Two interviewees also envisioned using English in their future careers. Four participated in overseas study

programs the previous summer and one attended the Summer Program in English sponsored by the affiliated university. In addition, one takes weekly conversation lessons at a language school.

Although English proficiency is their ultimate goal, it is not their most pressing one, especially for those who plan to take traditional entrance exams. Because success in Japan is often measured by the university from which one graduates, these students are hesitant to put all their eggs into the “English for communication basket,” so to speak. Like most college-bound students in Japan, they attend cram school to supplement the lessons they have at high school. They are aware that their cram school studies may not lead to communicative proficiency, but they nevertheless see it as essential to achieve their goals.

### 3.3.3 *The second interview*

The second interview took place in July 2017 shortly after all the students took the TEAP test and focused mainly on how the students prepared for the TEAP test, and their impressions after taking it.

At the time of this interview, only one student, Noriko, said she intended to use TEAP for university entrance and Megumi thought she might. The others, Kaori, Mariko, and Kimie, were going to take traditional entrance exams. The students all believed TEAP would help them improve their English skills, but as third year students who were under much academic pressure, they were unable to devote a great amount of time studying specifically for it. Kimie, a future science major, said she needed to concentrate not only on English, but also on science. Prior to TEAP, she was studying for science-related practice exams given weekly at her cram school. Megumi was also too busy with other academic subjects to narrow her focus down to just one test, as the following comment shows:

*I didn't study very much for TEAP. I didn't think about it consciously.*

*When I have classes at school, I have no room in my mind. In addition, I study at cram school once a week, and I don't have any time to study for TEAP. So I study hard for general comprehensive ability, that's not used only for TEAP, but can be used for TEAP.*

The students initially believed that TEAP would be similar to other proficiency tests they had taken, such as GTEC, EIKEN, and TOEFL iBT. They also believed that they would perform well on the reading and writing component because of their classes at the high school and at cram school. Even though they say they did not study much for the test, they did not want to do poorly on it so they bought self-study manuals several weeks before the test (and after the term exams were finished at the school) and practiced the exercises in them, took the mock tests, and checked their answers. After familiarizing themselves with the structure of the test, they focused on areas they believed they were weak in and/or areas that were different from the other tests. The following details how the interviewees studied for the individual components of the test.

#### *Speaking preparation*

The students were concerned over the speaking component, especially since their oral communication classes were reduced after entering the third year. Kimie said:

*After I became a third-year student, the classes at school were divided based on the level, and in my class, we don't have any chances for speaking. I didn't say a word of English, and we don't use a textbook, either. The handouts are distributed and translated, and I didn't have any chances for output. So, I felt terrible in this situation long before I took TEAP. So, I was speaking by myself in English based on the sample questions. But with this, there was no one who responds to me and I can do that by myself with my own pace. I was speaking slowly, thinking about how I should say that in English...English didn't come out quickly, but it was fun to speak English because I spoke for the first time in a while.*

The participants were particularly anxious about sustaining a one-minute monologue and playing the role of the interviewer during one section of the test, which was something they had never done before. Mari practiced interviewing her

cram school teacher. Megumi did the exercises in the TEAP book's CD to practice, but felt her timing was off. She practiced expressing her opinion on potential topics, trying to speak in a relaxed and confident manner. At lunchtime, she practiced with a friend, and when they were unsure of how to proceed, they checked Google.

Kaori used her smartphone to record herself, speaking for one minute every day for a week. Like Megumi, she wanted to evaluate her speaking speed. Kaori was unafraid of speaking in English because of the many opportunities given to her in her high school, despite feeling that she was not good at it. Noriko, on the other hand, was the only one who said that she practiced alone. This, she felt was unsuccessful because she was mostly "mumbling and not really speaking." Kimie believed that her speaking class had provided her with sufficient speaking practice. She also said that she spoke by herself. "I am a strange person, and when I was taking a bath, suddenly I start singing a song. I sing mainly foreign songs, but suddenly I think about various things by myself and just say them in English."

### *Listening preparation*

The students reported concentrating most on listening after realizing the TEAP listening passages were longer than what they were accustomed to. Also, they noted that unlike most educational materials published in Japan that exclusively use American voices, TEAP questions are read by speakers with a variety of accents. Thus, they tried to familiarize themselves with a range of accents, particularly through TED Talks. Kimie and Megumi realized their listening comprehension improved when they took notes, so they started practicing that. Kaori read aloud the listening script in the TEAP study book, believing that this would improve her listening skills.

### *Reading preparation*

The students felt most confident about the reading component, especially after being able to complete the practice tests within the allotted time. However, the TEAP readings were longer than those on other tests, so they began reading longer passages and tried to increase their reading speed. When they realized the content and the style was different from other tests, they looked up unfamiliar words related

to university life, such as *faculty*. Noriko said that after doing the practice tests, she checked the meaning of all the unfamiliar words. Finally, after Kaori answered a question incorrectly because she misinterpreted a graph, she began reading and interpreting more graphs.

### *Writing preparation*

The students were also somewhat confident about their writing skills because they had many opportunities to express their opinions in high school and cram school writing classes. Megumi and Kimie believed writing without a dictionary for thirty minutes as a regular activity in their high school English class would enable them to perform well on the test. After analyzing the sample essays in the practice book, they decided to concentrate on summarization and learn how to pick out the main points in the reading. Prior to the test, Kimie was working on an English newspaper project in school where she and her partner selected a topic and wrote articles in Japanese and in English based on the data. This, she said, was useful in preparing her for TEAP because she became accustomed to writing about data.

Megumi saw learning how to summarize as a valuable exercise, because she believed she would be required to write in such a manner after becoming a college student. She practiced writing and summarizing with a study partner and was able to take the TEAP test with confidence. Mari asked her cram school and high school teachers to evaluate her practice essays and to advise how she could improve it. Her high school teacher advised using direct expressions, paying attention to topic sentences, and giving reasons for her opinions. Based on his advice, she revised her essays. She also anticipated what kind of questions might appear on the test and she considered how she would answer them.

Noriko and Kaori, on the other hand, studied independently for the test by only comparing their answers with the sample answers.

### *Grammar and Vocabulary preparation*

None of the students said that they specifically studied grammar or vocabulary, but they checked their mistakes and looked up unknown words, as indicated above in

preparation for the reading test. Kaori mentioned it was easier to learn new words when she memorized entire sentences.

### *Reflections on the Test*

The four students who took the TEAP test perceived it to be much easier than the practice books and EIKEN Pre-1 level. The readings were challenging because of their length, but they were able to complete the reading and answer the questions with enough time to go back and double-check their answers. They also felt that testing of vocabulary, which is embedded in the reading test (Reading Part 1a), was easy because they could infer unknown meanings from the content.

The writing component was also relatively easy for them, thanks to the timed writing practice they had in school. They could finish with sufficient time for revision. Furthermore, in the summarization tasks, they could recycle words from the reading and switch them with synonyms that they knew. This was much easier than having to pull original words out of their head. Mari, however, had difficulties organizing her thoughts on whether she agreed or disagreed with the given topic.

As the participants anticipated, being asked to speak for one minute and act as an interviewer was challenging. Even though they had practiced beforehand, the test atmosphere made them nervous and their confidence slipped.

The most challenging part of the test was the listening component, namely because it required the test takers to make snap decisions. Unlike with the reading or the writing components, they could not think over their answers or go back and change them. Mari said there was one question she could not catch at all, and as a result, she was unable to answer.

### *3.3.4 The final interview*

The third and final interview was carried out in the spring of 2018, after the participants finalized their plans for the upcoming academic year. They focused on the entrance exams taken by participants and their current academic plans and the perceived usefulness of TEAP.

### *Participants' entrance exams and academic plan*



This section is particularly interesting as it provides a glimpse into the lives of five individual high school students' as they traverse the Japanese entrance exam landscape. As Table 6 shows, four of the participants took multiple exams for private, public, and national institutions. The numerous English exams and their various contents illustrate the situation facing high school leavers in Japan.

*Table 6: Summary of tests taken by interviewees and their immediate plans*

	English exams taken	What they will do
Megumi	Center Exam <i>Suisen</i> * at national university	Go to national university to study social studies
Mari	<i>Suisen</i> for pharmacy school Center Exam Regular exam	Go to private university to study pharmaceutical science
Noriko	Center Exam Private universities National university	Go to national university to study Chinese
Kimie	Center Exam National university Public university Private universities	Study one more year to enter medical school
Kaori	Center Exam National university Private universities	Study one more year to enter information science university

\*Special recommendation

Megumi obtained early admission to a national university through their special recommendation system (*suisen*), and so no further examinations were necessary for her. However, since she did not receive notification of her acceptance until mid-December in 2017, she continued to study for traditional entrance exams for the national university. This included not only English, but also other academic subjects necessary for the national Center exam. If she had been unable to gain admission to the national university, she would have used her TEAP scores to apply to private institutions instead. Even after receiving the offer from the university, she decided to take the Center Exam with her peers as a means to stay motivated to study.

Mari also took the early admissions examination for a pharmaceutical university, but failed. Therefore, she needed to follow the traditional examination route. Because her goal was to enter the science faculty of a university, she needed

to take the English examination that was written by the faculty she wished to enter in addition to the Center Exam. That exam consisted of long English readings followed by multiple-choice questions focusing on difficult grammatical points and vocabulary. The test also included one Japanese-English translation.

Noriko took entrance exams for both private and national universities. She was accepted into a private university, for which she used her TEAP scores. She continued studying for the national university entrance exam, and after she was accepted there, she turned down the offer from the private school. According to Noriko, the reading and listening components of that English exam shared some similarity with the TEAP test, but she also had to listen to a lecture and write her opinions about the lecture.

Unfortunately, Kimie and Kaori failed to enter the universities of their choice, and as a result, they chose to spend a year (or possibly more) attending preparatory schools dedicated to helping students pass entrance exams of high-level universities, particularly in fields such as medicine, arts, and sciences. Kimie sat the examinations for four medical schools. Despite her high score on TEAP (376/400), she found the English examinations at the various universities, all of which focused on different areas of English, to be challenging. She said this may have been because she devoted her study time to mathematics prior to the exam, which she said was her weakness. Nevertheless, the fact that Kimie could achieve such a high score on TEAP (CEFR B2), yet fail to achieve sufficient scores for entrance, is surprising and points to the variation in skills tested on English entrance exams. Most of these exams contained readings related to medicine followed by questions focusing on grammar and vocabulary. Some had Japanese-English and/or English-Japanese translations and one university required applicants to write an essay on lined paper. Since Kimie had heard the rumor that “they do not check it if it is less than half, [she] wrote as much as possible.”

Kaori wanted to enter a science faculty to study information science. Like Kimie, she needed to study for science-related subjects such as chemistry, physics, and math. And like the others, she was relatively confident of her English skills, thanks to her classes at the high school and at her cram school. The English component of the national university exam consisted of readings followed by

grammar and vocabulary questions, short translations, and one opinion essay. Like all of the other interviewees in this study, there was no speaking component of any of the exams she took. Needless to say, the lack of a speaking test means the university administrators cannot see how proficient these students had become in speaking, as demonstrated by their TEAP scores.

#### *Perceived usefulness of TEAP*

Because TEAP focuses on broader aspects of English, the interviewees felt the high scores they obtained demonstrated their actual English ability (as opposed to entrance exam English) and this motivated them to study ‘real’ English, especially after they completed their entrance exams. Furthermore, the participants believed that preparing for TEAP also prepared them for the type of English they would need after entering university. Megumi envisioned classes requiring her to synthesize the writings of others and not just express her own opinions on general topics. Therefore, she felt that learning how to summarize would be a useful skill for her future studies that is not focused on in other tests. This notion is reflected in the following comment: “Writing includes reading [in TEAP]. In GTEC, [the exam question is] written in Japanese, and I have to write, so I need my own imagination and my English ability is challenged. But in TEAP, summarizing is important. Probably that’s what we need for university—not only summarizing my own opinions, but trying to find appropriate parts and understand others’ [opinions].”

In addition to improved linguistic ability, the participants said studying for TEAP prepared them for various aspects of the other entrance exams, particularly when it came to reading and/or listening to long passages and to writing essays in a limited amount of time. Noriko, for example, said that studying for and taking TEAP was good preparation for some of the other entrance exams. Some universities tested applicants’ listening and writing skills in a way similar to TEAP, and therefore, after receiving scores on those components of the TEAP test, she adjusted her study methods. This preparation enabled her to get good scores and enter the university of her choice. Kaori also felt that the preparation for the writing portion of the TEAP test came in handy on some of her other examinations. She had been required to write an essay on a provided sheet of A4 paper in one exam. She was able to “write

as much as possible” thanks to being accustomed to writing long pieces in her classes and for TEAP.

Although participants thought preparation for TEAP was beneficial, a combination of other factors enabled them to do well on the English university examinations they took. Most importantly, these students were highly motivated, ambitious, and studious. They reported approaching other proficiency tests (e.g., G-TEC, TOEFL) with the same purpose, that is, wanting to get high scores that showed improvement in their English. In addition, the teachers in their high school focused not only on *juken*-focused language instruction, but also on how to improve general English skills.

A final point was that some students felt the topics related to student life in a university, such as applying for scholarships, were somewhat unfamiliar to them and different to other tests they had studied for. As such, they felt that the skills and knowledge required for dealing with these topics may not be transferable to other tests. Overall, however, all participants felt that studying for TEAP was useful and practical, regardless of whether or not they could use their score for admission purposes.

#### *3.3.4 Teacher Interview*

To help clarify and consolidate our understanding of the learners’ test, survey and interview data, we interviewed their high school English teacher. The teacher has been working in the particular high school for a considerable amount of time and is very familiar with the students and their studies. Although the students undertake lessons with other English teachers as part of the curriculum, this teacher oversaw their learning during their second year, which is when the majority of this study took place. He was therefore able to provide additional information about the in-class activities as well as typical homework assignments. The interview was conducted as an informal, free format discussion regarding the students’ learning. We asked questions that particularly focused on the distribution of skills work in classes and for homework and the types of activities that students did, as well as picking up some of the issues that students had raised during the interviews and the general findings of the test and survey data.

From the interview the following conclusions can be drawn. Firstly, the teacher strove to give students a rounded English language learning experience during their second year by balancing the four skills as far as possible. However, he tended to focus on speaking activities in class as he believed this was something students would benefit most from during class time, whereas other skills work could be done outside of class. The teacher encouraged speaking class through using English himself during his classes. Speaking activities included role-plays, prepared and spontaneous monologues given by students in small groups followed by a number of the more confident students who gave their monologues to the whole class, and small group discussions. He believed in the importance of giving students some topic to discuss in the form of pre-reading and pre-listening activities, rather than asking students simply to discuss things related to their daily lives (e.g., 'what they had for dinner last night'). Readings and listening texts were taken from various sources including the class textbook and other textbooks and test (e.g., TEAP, TOEFL, GTEC, EIKEN, IELTS) preparation materials, and from online resources, such as TED Talks and news channels. He saw benefit in using jigsaw activities where students are given different sources of information (including texts, graphs, pictures) and needed to combine these in some way, usually beginning with an oral summary of the information. For role-play activities, he saw benefit in giving sufficient role information, which helped the learners become more involved in the interactions. He mentioned that he liked the TEAP Part 2 speaking task, in which the test-taker interviews the examiner. Especially, he saw this as important because it made students ask direct questions, which is something that they need to do but tend not to do, perhaps due to cultural differences between English and Japanese. All in all, it is clear that his approach was geared towards having students use English in class in the oral mode.

His approach to writing was also, in our opinion, pedagogically sound, and he encouraged writing on various topics and using various methods. For example, all students were given an A4 notebook at the beginning of term and were asked to write about anything they like in it. Observation of these note-books revealed that students were diligent in completing the task and wrote about a wide variety of topics related to their interests and in various formats: for example, notes taken

from TED Talks which were then developed into a summary text followed by an opinion text; short informal essays detailing recent trips; detailed descriptions of baseball teams and the students' preferred starting line-ups; and so on. Students were encouraged to share their favorite writings with other students to encourage learning of various writing styles. The teacher noted, however, that most writing was done at home, with in-class activities focusing on either feedback or discussion of the topics. This was seen as a way to maximize students' productive use of language in class. His teaching involved both summary writing, which he saw as a very common type of activity used in Japanese high school classes, and free writing, which he noted his students were able to accomplish due to their relatively high level of English writing ability. (In other high schools, he saw the initial and major difficulty for students being actually writing anything up to a paragraph in length). He encouraged paraphrasing and simplifying the source text as necessary to communicate the main ideas accurately. As mentioned previously, he took inspiration from a range of test preparation materials and this was true of his approach to task design, for instance incorporating both summary writing and free writing in class. In sum, the teaching approach to writing appears to accommodate students' interests as well as supporting their skill development.

Regarding reading and listening activities, the teacher encouraged reading and listening to different types of texts, including academic texts, both during and outside of class time. He recognized the utility of using online resources and often had students listen to part of a text in class and then provided access to the full text, which students could access on their smartphones during break times or at home. He encouraged note-taking of materials followed by summary and opinion writing and/or speaking, in order to develop note-taking skills, which he understood as an important skill that students needed to develop, particularly for the university context. He recognized that his students tended to prefer to listen to TV, music, movies and talks in English, because they were interesting and easily available online. He also noted the difficulty in selecting materials due to the varied proficiency levels of his students: many of the texts were simultaneously just right or too difficult/easy for different students in the class. It is clear from the teacher's perspective that reading and listening are taught in an integrated fashion with

speaking and writing. That is, rather than passively reading and listening, students need to comprehend in order to use this information for a subsequent communicative task.

## 4. Discussion

In this section we bring together the findings of the test, survey and interview data and provide answers to the research questions.

### 4.1 Research question 1

What were the scores for the TEAP Test and for each component skill? Was there a bias towards receptive skills as seen in other studies with other tests in the Japanese context?

TEAP distinguishes between proficiency ranges from CEFR A2 to B2 (with a top score of 100 being equivalent to C1, based on 2017 data), because this is the typical range of scores for high school leavers who apply to university in Japan. The students in the present study were at the higher end of this range: survey participants were generally at the B1 level and the interviewees were at the B2 level. This may be considered a reasonably high level of proficiency given that participants had just entered their third year of high school.

Scores were higher on productive skills with both writing and speaking being over ten points higher than reading and listening. However, these scores cannot be interpreted as evidence that the students were higher proficiency in productive skills than in receptive skills. According to the benchmarks for TEAP scores and the CEFR scale (of 2017), the B1 level includes the following scores: Listening: 51~74; Reading: 50~74; Writing and Speaking: 61~84. This means that although participants scored higher in productive skills, their scores were nevertheless in the middle of the score range for the B1 level. It is also noteworthy that the interviewees typically scored considerably higher for speaking than the other participants. In sum, the participants in this study are higher proficiency learners of English and are equally strong in productive skills and receptive skills.

These results are markedly different from previous studies in which scores tended to be much higher for receptive skills (Allen, 2017). They apparently show that participants had not yet experienced washback from entrance exams which encourages a focus on receptive skills over productive skills. This may be because the students were just entering their third-year of high school when they took the test



and entrance exam preparation typically ramps up during the third year. The interviewees support this assumption through their comments about doing less work on productive skills in their third-year.

#### *4.2 Research questions 2 and 3*

Do test takers' preparation strategies differ from those that they normally employ when learning English? If so, does the change indicate positive washback on learning processes?

What factors mediate the process of washback from the TEAP Test?

The following factors mediated washback to the test in general and thus are discussed prior to focusing on washback to skills and sub-skills.

- **Test importance:** According to survey data, the majority of participants were unsure whether the TEAP score would be useful for their future and more thought that it would not be useful than those who thought it would. Overall, in terms of the utility of the test score, which is undoubtedly a factor in perceived test importance, it is likely that most participants did not perceive the test as high-stakes. Consequently, the amount of preparation students did for the test, and thus the amount of washback, was limited by this factor. Most participants reported studying less than 20 hours for the test (78% of participants), which underscores the overall impact of the test on learners' behavior.
- **Test difficulty:** Participants perceived the test to be challenging but achievable. Notably, participants perceived the writing section of TEAP to be less difficult than they had anticipated, as indicated by comparing the data from the two surveys. According to interviewees' responses, this was due to the provision of texts from which vocabulary could be paraphrased; for the learners, this task was perceived as easier than the freer writing tasks where only the question prompt is provided. It should also be noted that students had had plenty of experience of summary writing during their high school classes.
- **Expectancy of success:** Participants expected that they could succeed on the test if they prepared.
- **Motivation:** Participants were highly motivated to study all skills both before and

after the test. They considered English to be very important for their future and their perceived importance of skills mirrored their reported focus on skills in high school classes (i.e., speaking was studied the most and viewed as most important, while writing was studied the least and was viewed as least important). While this could be coincidental, it is likely that participants were influenced by the content of their high school syllabus and their teacher's approach. Also, from the interviews it was clear that at least some of the students like the challenge of taking a test, even if it was not crucial for their future.

Taken together these mediating factors are important for determining the strength of the washback effect. While motivation, expectancy of success and test difficulty were all likely to promote a strong washback effect, the lack of importance of the test for many test-takers likely undermined this effect. That is, even though many of the conditions were appropriate to engineer test washback, if test-takers do not perceive the test as important, the washback effect will be greatly reduced.

#### *Focus on skills*

Overall, according to the survey data, participants reported studying reading and writing more than listening and speaking both during their test preparation and independent study (*Independent: R>W>L>S; Test: R>W>L>S*). Speaking was the least studied skill for the test as well as independent study, and although participants reported studying speaking slightly more for the test than during general independent study, the difference was not significant. In other words, while there was an increase in the amount of out-of-class study focusing on speaking, and this was due to the test, the change was not statistically reliable.

One of the intended consequences of TEAP is the engineering of positive washback on speaking skills. However, this was not reliably demonstrated in this study. Perhaps the main reason for the lack of washback on speaking (i.e., an increase in the amount of study and practice of speaking) is the fact that students reported studying speaking most in class and their scores were high overall for speaking. Therefore, students had plenty of opportunity to practice speaking in class

and this led to high scores on the speaking component. Interviewees stated that they were quite confident about speaking, though their confidence dipped by the time they took the test because the number of classes in which they practiced speaking had greatly reduced after entering the third year. Finally, the teacher interview data supports the view that speaking was a primary focus of in-class activities. In sum, within the micro-context of the school, the focus on productive skills in class may be the primary reason why greater washback effect was not observed on studying speaking.

Another reason why there was limited washback on speaking was that participants appeared to strike a balance between in-class and out-of-class study: the focus of high school classes tended to be communicatively oriented in the oral mode ( $S > L > R > W$ ), while the focus of independent study was typically in the written mode ( $R > W > L > S$ ). Participants were preparing independently for the test (i.e., in a non-instructed context) and therefore test preparation was part of their independent study. Given that their focus of independent study was more geared towards reading and writing, it appears that they carried this trend over to their test preparation study. This could be justified if learners believed they had sufficient and relevant practice of speaking and listening in school.

The role of internal and external resources (Green, 2007) are also important to consider regarding the lack of washback on speaking. Internal resources concern cognitive abilities, knowledge and experience; for example, whether participants know how to study a particular skill, such as speaking. The findings of Allen (2016a, 2017) revealed that Japanese undergraduates often did not have much experience of studying speaking, in general or for tests, and therefore were at a loss as to how to approach preparation for the IELTS speaking test. In the present study, in contrast, participants appeared accustomed to speaking at school and therefore had a good idea of how to practice speaking in English for the TEAP test. They also had had experience of taking other tests, both English-related and others, which means they were familiar with the general skills required for test preparation (i.e., time-keeping, organizing study time). In sum, there was no clear lack of internal resources in the present findings that could explain the lack preparation for the speaking test.

Regarding external resources, these may be material (e.g., books, online

resources) or human (i.e., peers, teachers or native speakers from whom to gain feedback and practice using English). The data demonstrate that learners had access to test-related materials and other learning materials and facilities in school. Therefore, it is unlikely that washback was impeded by lack of material resources. In contrast, participants generally did not employ social strategies (i.e., seeking advice, feedback and conversation partners) when preparing for the test. This may be due to the unavailability of such resources: In Allen (2016a, 2017), participants often reported that because they had no one to speak to, they did not practice speaking. That is, they believed that a conversation partner, or someone able to provide feedback, was essential for practicing, and learning, speaking. However, such comments were generally lacking in the present study; conversely, interviewees reported using social strategies such as asking their teacher and cram school instructors for help, as well as practicing with one another and at conversation school. In other words, they knew where to get help when they needed it. The interviewees were higher proficiency and may therefore be more experienced in studying English, especially English speaking, which would explain this discrepancy between the mean agreement ratings on the survey and the comments made by interviewees. Overall, however, we are inclined to believe that participants may have had access to external resources but were not especially inclined to use them. This was likely because they were reasonably confident from practicing in school, and also because they did not value the test as critically important and therefore did not spend much time preparing for it. In a different situation, where the test is important and students do not practice speaking/writing a lot in class, we would expect the role of external resources to be much more important. We believe this is a crucial issue for future washback research with TEAP.

Writing was the second-most focused-on skill in both independent and TEAP-focused study, but the least focused-on skill in school. Nevertheless, participants also scored high on writing. They also reported that writing was the easiest section of the test. In other words, even though learners reportedly did not focus on writing so often in class, they nevertheless achieved a high level of writing ability. This discrepancy is likely due to the fact that writing was typically assigned as homework, according to the teacher, which concurs with reported focus on reading and writing

during out-of-class study.

The general finding that students focused mainly on speaking and listening in class but reading and writing at home appears somewhat at odds with the broader Japanese context. In the data provided by Green (2014), third-year high-school students reported that in-class activities in Japanese high schools (n=75) were heavily weighted towards written language, with two thirds of classes paying attention to listening skills, and only between a fifth and quarter of classes devoting time to speaking (p.16). Thus, the present high school context clearly falls into the minority of high schools that devote most time to study in the oral mode. However, the participants in Green's study were third-year students and, as mentioned by interviewees, the focus on the written mode was expected to increase in their third year, which may explain some of the discrepancy. Nevertheless, the present study context is clearly part of a minority of schools that provide a balanced approach to teaching and learning of the four skills.

#### *Focus on task-specific skills*

Evidence of positive washback was notable in the following areas:

- **Increased reading and writing under time pressure:** Due to the task demands, which include reading and writing long texts, participants adjusted their test preparation accordingly. However, although practicing reading and writing under time pressure is likely to positively affect the development of fluency in language use, it is not particularly unique to TEAP: test-takers of most exams will prepare by practicing written-mode responses under time pressure.
- **Interpreting and writing about visual data and synthesizing information:** These skills are ultimately important in the university context and demonstrate that task-specific washback does occur in line with the intended washback outlined by the test designers (Green, 2014, p.38). This novel feature of TEAP, in comparison to most Japanese entrance exams, thus guides learners to develop critical thinking skills and higher-level writing abilities that are required in the TLU environment (i.e., academic situations) but also extend to other areas of the real world (e.g., business). Even though the teacher interview data suggests the use of visual materials in class, for example in jigsaw activities that involved both

writing and speaking about data, was common, the TEAP test nevertheless appeared to stimulate further practice in these sub-skills.

- **Listening to a variety of accents:** One interviewee reported listening to a range of speakers to deal with the task demands of the test; this is likely to benefit her in situations where she has to comprehend a range of spoken English varieties. Although this is not explicitly mentioned in the intended washback for listening (Green, 2014, p. 37), it can certainly be considered positive washback based on perceptions of task demands.
- **Listening to a variety of genres:** Interviewees listened to a wider range of genres to develop their listening ability. This is in line with the intended washback and will certainly benefit learners who wish to use English in various situations, both for personal and work/academic purposes.
- **Checking vocabulary related to academic situations:** One interviewee noted doing this in response to the requirements of the test. Developing knowledge of academic vocabulary is important for using English on university campuses and this is explicitly stated as intended washback (Green, 2014, p.37).

Additionally, evidence of positive but indirect washback was found in open-ended survey responses, which revealed a wide range of behaviors (including those mentioned above) that participants reportedly adopted following the preparation sessions. This can be considered washback from the test, via the information provided and activities conducted in the preparation sessions. As can be seen in Appendix 3, there are many cases of individuals adapting their behavior to meet the (perceived) test demands, and almost all of these can be considered positive changes in learning behavior.

Evidence of ambiguous washback was found on the following:

- **Less reading of fiction:** Participants read less fiction in preparation for the test than in their general English study. Participants obviously realized the lack of fictional texts in TEAP, which includes more academic and campus-related texts, and changed their learning behavior accordingly. However, whether this can be

called positive washback is unclear because reading for pleasure is positively associated with development of reading fluency and proficiency in general (e.g., see Day & Bamford, 1998).

- **Less note-taking while listening:** Participants reported practicing note-taking significantly less for TEAP than in general. Note-taking is notably absent from the intended washback specifications (p.37), and it is therefore unclear whether participants are expected to develop (or even utilize) this ability when preparing for, and taking, TEAP. Undoubtedly, note-taking while listening to long monologues, such as lectures, is one of the most fundamental skills required at university and thus forms part of the construct of listening ability in the academic context. However, although TEAP includes such long monologues in Listening Parts 2A and 2B, it is unclear whether note-taking is necessary for success on the tasks. In fact, one interviewee mentioned that taking notes impeded her ability to select the correct answer during the test and therefore decided not to take notes. Therefore, it may be that participants perceived note-taking as unnecessary or even disadvantageous for the test, and thus did not practice it. Alternatively, the teacher interview data suggests learners were relatively accustomed to note-taking of audio texts, which may explain why they did not practice significantly for the test. While we cannot conclude decisively on this issue, the role of note-taking in TEAP listening perhaps requires further consideration from the perspectives of both construct and consequential validity.
- **Less practicing of role-plays:** Role-playing was also studied significantly less for the TEAP test than in general, in-class English study, even though role-playing is fundamental to the speaking test as a whole, but particularly Part 2 (where the test-taker becomes the interviewer). This specific finding may be understood in relation to the broader finding that speaking is primary in classwork. In other words, because students are familiar with role-plays from their classwork, they did not see the need to study it particularly for the test. Therefore, while participants may have identified the need to perform well in English role-plays for success on the test, they felt that they had prepared adequately for it in class activities. Thus, washback on this sub-skill was absent because it was already part of the class syllabus. Alternatively, because the learners prepared for TEAP

independently, and primarily at home, they may have had difficulty in finding an interlocutor with whom to perform the role-play. This may also explain why learners were reluctant to practice role-playing during their test preparation. In any case, because role-playing is specified in the intended washback as a sub-skill to be developed by test-takers, the TEAP test designers may need to consider ways in which they can promote washback to this specific ability. For instance, providing materials and examples of *how* to practice by oneself for this section of the test may be of great use to learners and thus promote development of oral role-playing ability when access to interlocutors is limited.

No evidence of washback was observed for the following:

- **Reading academic texts**
- **Reading and listening to information about universities**
- **Listening to texts while attending to visual data**
- **Speaking about more abstract topics**
- **Using more complex grammar and vocabulary in speaking.**

These are aspects of the test that were neither studied much before or after introduction of the test (i.e., had ratings of  $M < 3.0$  on both surveys), although they are important aspects of TEAP and intended as target skills for intended washback. In other words, while doing all of these activities is considered important for success on the test, participants did not do them any more after the introduction of the test.

A general explanation is simply that there was not enough time for participants to focus on many different aspects of test preparation. Participants did not spend a great deal of time preparing for the test during their independent study (36 out of 46 studied  $< 20$ hrs, while 10 studied  $> 20$ hrs). Therefore, given the limited amount of time spent preparing for the test, participants were limited in what they could attend to during test preparation.

Also, a specific explanation for not studying information about universities can be proposed based on the interview data. Interviewees mentioned that while they realized it was important for success on the test, the focus on English-medium campus-life seemed irrelevant to their needs because they would be attending



university in Japan, where Japanese is the primary language used on campus. If students do not perceive this to be important, they may be less inclined to focus on it, even if it is perceived as important for the test. This is an issue that should be considered in future studies into learner perceptions of TEAP.

Regarding reading academic texts, given that learners performed relatively less well on the reading test, it may be that this is something that they should have focused on more. This may again be related to context: interviewees explained how classes had focused more on communication in the second year but would shift to a more written-mode, *juken*-style of English education in the final year. This style of learning tends to focus on reading academic texts, translating parts of them, and doing vocabulary and grammar-related exercises based on them. Thus, strategically speaking, learners may have been reluctant to start reading additional academic texts prior to the test because they were aware they would have to do so thereafter. Alternatively, this could point to participants' interpretation of the survey item (i.e., 'academic texts' could have been interpreted as academic articles or the like), given that they reported practicing reading academic texts (i.e., those in TEAP) under time-pressure for the test.

#### 4.3 Research question 4

What were test takers' views on the TEAP Test and their experience of preparing for it? Did they feel the test helped them to improve their English?

Participants' comments suggest that they see the value in the tasks required by TEAP, and this may have reinforced and/or changed their beliefs about how to study English for authentic purposes rather than *juken* purposes, which require them to acquire skills that will be of more use to them in their long-term English study goals. In particular, they recognized the tasks to be similar to what would be required of them as university students: synthesizing information from reading, understanding visual aids such as graphs, and writing about academic topics. They felt that learning unknown vocabulary through the contexts was holistic and integrated.

Participants were also broadly in favor of having speaking tests on entrance exams (75% agreeing to some extent), again illustrating their perceived importance

of being able to speak English, as well as their desire to have their speaking ability assessed. This compares favorably with views of a much larger sample of high school students ( $n=3,868$ ) reported in Green (2014), 62.4% of whom, according to 6-point Likert scale, agreed to some extent with the necessity of speaking tests. Again, this illustrates how positively students viewed speaking skills, reflecting their study of speaking and achievement on the speaking part in the TEAP test. Nevertheless, it should be remembered that around 25% of those in the present study, and closer to 37% in Green (2014), were against the use of speaking tests for entrance purposes. Further consideration of this hostility and concern is imperative for test developers aiming to generate positive washback in Japan.

Participants perceived their actual improvement from studying for the test as significantly *less* than their predicted improvement. In other words, while they thought that preparing for TEAP would help them to improve their English, they were less convinced that they had improved after preparation. This result needs to be understood in context, particularly regarding the perceived importance and the amount of preparation that participants actually did for the test. It is likely that participants were initially motivated to study hard for the test but this motivation was re-directed to other subject areas over the course of the study. One reason for the re-direction of effort was because the students realized closer to the time of the test that they were unlikely to need the TEAP score for entrance purposes. Therefore, they directed their efforts to subjects, and tests, that would help them achieve their proximal goal of university entrance.

## 5. Conclusions

The TEAP test brings with it a vision of promoting positive washback in Japanese high school in terms of the skills and sub-skills learners develop through their English education and test taking experiences. In the present study, we found strong evidence of positive washback on specific sub-skills relating to academic reading and writing. Learners' responses to open-ended questions in the survey and interviews also indicated various specific instances of positive washback. Moreover, participants were in favor of speaking tests and viewed the TEAP test positively. They drew a sharp distinction between the attributes of TEAP and those of other entrance exams: TEAP was believed to provide a thorough assessment of their communicative English ability while other entrance exams were seen as important for their immediate goals, but not for developing communicative abilities in English. These findings demonstrate the potential for TEAP to have positive impact on high-school learners' English development.

The positive impact of TEAP in the present study was, however, limited due to certain mediating factors. Firstly, most of the students in this study did not need TEAP scores for entrance to university because universities they applied did not require TEAP score for admission. As a result, the amount of preparation was restricted and instead students' efforts were directed to other studies and test preparation. Secondly, the students were fortunate in that they received instruction and opportunities aimed at developing their productive abilities in school classes. Consequently, they did not feel the need to increase the amount of study of these skills specifically for the TEAP test. This, however, is not the norm in Japan and much greater washback is expected on productive skills particularly in situations where school-based instruction is largely aimed at developing receptive abilities, which are typically tested on university entrance exams.

Overall, this study has provided data that will help educators and assessment practitioners understand the complexity of washback and the range of factors that mediate its intensity and direction. The data particularly reflects the study behaviors of motivated students of higher proficiency in English and therefore does not necessarily generalize to the majority of high school language learners in Japan.

Nevertheless, it is clear that TEAP has the potential to create positive washback on learning. Future studies should seek to further confirm the consequential validity of the test from the point of view of washback on learning. Such studies will contribute to the ongoing validation and development of the TEAP test.

## 6. Recommendations

In general, given the limited scale and single school context of the present study further research is needed into learning and test-preparation behaviors **on** a larger and more representative scale so that washback from TEAP can be evaluated

- 1) at a broader range of English proficiencies,
- 2) in schools where more traditional methods (i.e., a focus on written mode) are typically employed,
- 3) and with a larger number of participants who require TEAP scores for entrance purposes.

Additionally, based on the findings, the following specific points also deserve attention from test designers and researchers:

- The impact on learning by tests such as TEAP in terms of directing learners away from fictional texts and towards non-fictional, academic texts: Is it beneficial to exclude fictional texts from tests if they reduce high school learners' desire to read them?
- The role and importance of note-taking in TEAP listening: Is note-taking required for the test? Does the test promote the development of this important sub-skill?
- What are the ways to increase awareness of the importance of listening to different varieties of spoken English and a range of spoken genres? What are the ways to achieve this in practice (i.e., the resources available)?
- The rationale for including campus-based English dialogues in reading and listening activities: How can learners be convinced of the need to learn language for such situations? How can they prepare for this?
- The kind of support provided to learners by test and materials developers in terms of how learners can practice role-plays: How can learners practice role-plays if they have limited opportunities to do so in class or with other interlocutors?
- Is the writing component, which provides texts from which ideas and language are provided, indeed perceived as 'easier' than other writing tests? What is the potential impact of this perception on learning?

## 7. References

- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115–129.
- Allen, D. (2016a). Investigating washback to the learner from the IELTS test in the Japanese tertiary context. *Language Testing in Asia*, 6(7), 1-20. DOI 10.1186/s40468-016-0030-z
- Allen, D. (2016b). Japanese cram schools and entrance exam washback. *Asian Journal of Applied Linguistics*, 3(1), 54–67.
- Allen, D. (2017). Investigating Japanese undergraduates' English language proficiency with IELTS: Predicting factors and washback. *IELTS Partnership Research Papers* 2. IELTS Partners: British Council, Cambridge English Language Assessment and IDP: IELTS Australia. Available at <https://www.ielts.org/teaching-and-research/research-reports>
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257–279.
- Bourdieu, P. (1990). *In other words: Essays towards a reflexive sociology*. Cambridge: Polity Press.
- Brown, J. D., & Yamashita, S. O. (1995). English language entrance examinations at Japanese universities: What do we know about them? *JALT Journal*, 17, 7-30.
- Cheng, L. (2005). *Changing Language Teaching through Language Testing: A washback study (Studies in Language Testing 21)*. Cambridge: Cambridge ESOL/Cambridge University Press.
- Cheng, L. (2014). Consequences, Impact and Washback. In A. J. Kunnan (Ed.), *The companion to language assessment: Evaluation, Methodology and Interdisciplinary themes* (pp. 1130–1145). Singapore: Wiley Blackwell.
- Cheng, L., Andrews, A., & Yu, Y. (2011). Impact and consequences of school-based assessment (SBA): Students' and parents' views of SBA in Hong Kong. *Language Testing*, 28(2), 221–249.
- Day, R. R., & Bamford, J. (1998). *Extensive reading in the second language classroom*. Cambridge: Cambridge University Press.

- Gosa, C. M. C. (2004). *Investigating Washback: A Case Study Using Student Diaries*. Lancaster: Unpublished PhD thesis, Department of Linguistics and Modern English Language, Lancaster University.
- Green, A. (2005). EAP study recommendations and score gains on the IELTS Academic Writing test. *Assessing Writing*, 10, 44–60.
- Green, A. (2006). Washback to the learner: Learner and teacher perspectives on IELTS preparation course expectations and outcomes. *Assessing Writing*, 11, 113–134.
- Green, A. (2007). *IELTS Washback in Context: Preparation for academic writing in higher education (Studies in Language Testing 25)*. Cambridge: Cambridge University Press.
- Green, A. (2014). *The Test of English for Academic Purposes (TEAP) Impact Study: Report 1 - Preliminary Questionnaires to Japanese High School Students and Teachers*. Tokyo: Eiken Foundation of Japan.
- Hughes, A. (2003). *Testing for language teachers*. Cambridge: Cambridge University Press.
- In'nami, Y., Koizumi, R. & Nakamura, K. (2016). Factor structure of the Test of English for Academic Purposes (TEAP®) test in relation to the TOEFL iBT® test. *Language testing in Asia*, 6(3), doi:10.1186/s40468-016-0025-9
- Kikuchi, K. (2006). 'Revisiting English Entrance Examinations at Japanese Universities After a Decade', *JALT Journal* 27(1), 77-96.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13, 241–256.
- Mickan, P., & Motteram, J. (2009). The preparation practices of IELTS candidates: Case studies. In J. Osborne (Ed.), *IELTS Research Reports* (Vol. 10, pp. 223-262). Canberra: IELTS Australia and Manchester: British Council.
- Nakatsuhara, F. (2014). *A Research Report on the Development of the Test of English for Academic Purposes (TEAP) Speaking Test for Japanese University Entrants – Study 1 & Study 2*. Tokyo: Eiken Foundation of Japan.
- Nakatsuhara, F., Joyce, D., & Fouts, T. (2014). *A Research Report on the Development of the Test of English for Academic Purposes (TEAP) Speaking Test for*

- Japanese University Entrants – Study 3 & Study 4*. Tokyo: Eiken Foundation of Japan.
- O’Sullivan, B., & Weir, C. J. (2011). Test development and validation. In B. O’Sullivan (Ed.), *Language Testing: Theories and practices* (pp. 13–32). Basingstoke: Palgrave Macmillan.
- Saville, N. D. (2010). Developing a model for investigating the impact of language assessment. *Research Notes*, 42, 2-8.
- Saville, N. D., & Hawkey, R. (2004). The IELTS impact study: Investigating washback on teaching materials. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 73–96). Mahwah: Lawrence Erlbaum Associates.
- Shih, C. M. (2007). A new washback model of students’ learning. *Canadian Modern Language Review*, 64(1), 135–162.
- Smyth, E., & Banks, J. (2012). High stakes testing and student perspectives on teaching and learning in the Republic of Ireland. *Educational Assessment, Evaluation and Accountability*, 24(4), 283–306.
- Stoneman, B. W. (2006). *The impact of an exit English test on Hong Kong undergraduates: A study investigating the effects of test status on students’ test preparation behaviors* (Unpublished PhD thesis). Hong Kong: The Hong Kong Polytechnic University.
- Taylor, L. (2014). *A Report on the Review of Test Specifications for the Reading and Listening Papers of the Test of English for Academic Purposes (TEAP) for Japanese University Entrants*. Tokyo: Eiken Foundation of Japan.
- Tsai, Y., & Tsou, C. (2009). A standardized English Language Proficiency test as the graduation benchmark: Student perspectives on its application in higher education. *Assessment in Education: Principles, Policy & Practice*, 16(3), 319–330.
- Tsagari, D. (2007). *Investigating the Washback Effect of a High-Stakes EFL Exam in the Greek Context: Participants’ Perceptions, Material Design and Classroom Applications*. UK: PhD Thesis, Submitted in Department of Linguistics and English Language, Lancaster University.



- Watanabe, Y. (2004). Teacher factors mediating washback. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Context and method in washback research: The influence of language testing on teaching and learning* (pp. 129–146). Hillsdale: Lawrence Erlbaum.
- Wall, D., & Horak, 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. *TOEFL iBT® Research Report, TOEFL iBT-17*.
- Weir, C. J. (2005). *Language Testing And Validation: An Evidence-based Approach*. London: Palgrave-Macmillan.
- Weir, C.J. (2014). *A Research Report on the Development of the Test of English for Academic Purposes (TEAP) Writing Test for Japanese University Entrants*. Tokyo: Eiken Foundation of Japan.
- Weir, C. J., Vidaković, I., & Galaczi, E. D. (2013). *Measured constructs: A history of Cambridge English language examinations 1913–2012 (Studies in Language Testing 37)*. Cambridge: Cambridge ESOL/Cambridge University Press.
- Xie, Q. (2013). Does Test Preparation Work? Implications for Score Validity. *Language Assessment Quarterly*, 10(2), 196–218.
- Xie, Q., & Andrews, S. (2012). Do test design and uses influence test preparation? Testing a model of washback with structural equation modeling. *Language Testing*, 30(1), 49–70.
- Zhan, Y., & Andrews, S. (2014). Washback effects from a high-stakes examination on out-of-class English learning: insights from possible self theories. *Assessment in Education: Principles, Policy & Practice*, 21(1), 71–89.
- Zhan, Y. & Wan, Z. H. (2016). Test Takers' Beliefs and Experiences of a High-stakes Computer-based English Listening and Speaking Test. *RELC Journal (pre-print online version)*. doi:10.1177/0033688216631174

### *Acknowledgements*

This study was funded by the Eiken Foundation of Japan. We are particularly grateful to Keita Nakamura and Kazutaka Tanimoto at the Eiken Foundation of Japan for their generous support and assistance throughout the project. We would also like to thank Takayuki Tsukui and other members of the high school for their enthusiasm and kind support. Finally, we are grateful to Saki Suemori for her contributions as research assistant and to Emeritus Professor Edward Schaefer for reading a previous version of this report.

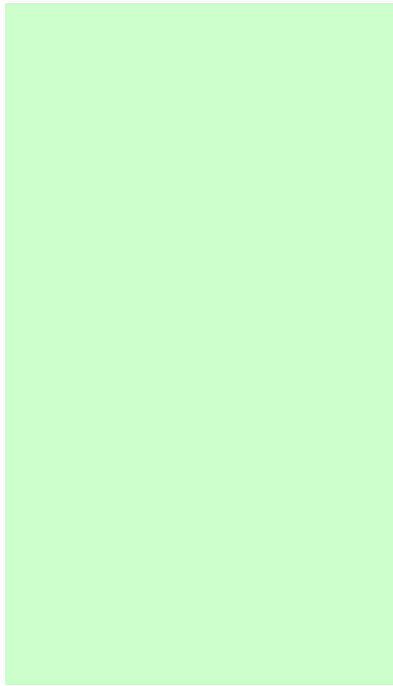
## 8. Appendices

### Appendix 1: Survey Items

Pre-Test Survey (Survey 1)			Post-Test Survey (Survey 2)		
Topic	Item	Response	Topic	Item	Response
	<b>English learning</b>				
Learning history	<i>When did you start learning English?</i>	0-2,3-5,6-11, 12-15, 16+			
Learning history	<i>Have you lived in an English-speaking country (including study abroad)</i>	Yes, No			
	<i>– if so, for how long?</i>	Open			
Learning history	<i>How often do you speak in English outside of class?</i>	Every day, almost every day, often, not very much, not at all			
Perceived Proficiency	<i>Rate your English language proficiency in the four skills (scale 1-10)</i>	0-9	Perceived Proficiency	<i>Rate your English language proficiency in the four skills (scale 1-10)</i>	0-9
Time	<i>How many hours on average per week do you spend studying English?</i>	1hr or less, 2, 4, 6, 8, 10, more than 10	Time	<i>How many hours in total did you spend studying English for the test?</i>	0,<20, 20-40, 40-60, 60-80, 80-100, 100+
Amount	<i>I spend a lot of time studying R/W/L/S in HS classes (including class homework)</i>	Likert x 4	Amount	<i>I spend a lot of time studying R/W/L/S for the test</i>	Likert x 4
Amount	<i>I spend a lot of time studying R/W/L/S independently (by myself, or at cram/conversation school, extracurricular activities)</i>	Likert x 4			
Motivation	<i>English is important for my future</i>	Likert			
Motivation	<i>English R/W/S/L ability is important for my future</i>	Likert x 4			
	<b>Perceptions of TEAP</b>			<b>Reflections on TEAP</b>	
Test Experience	<i>Have you taken the TEAP Test before?</i>	Yes, No			
Test Use	<i>I am planning to use my TEAP score for university admittance</i>	Yes, No, Not sure yet	Test Use	<i>I am planning to use my TEAP score for university admittance</i>	Yes, No, Not sure yet
Test Value	<i>My TEAP Score will be useful for my future</i>	Likert	Test Value	<i>My TEAP Score will be useful for my future</i>	Likert
Test Value	<i>I believe that preparing for TEAP will help me improve my English</i>	Likert	Test Value	<i>Studying for TEAP helped me to improve my English</i>	Likert
Expectation of success	<i>If I prepare well, I think I can do well on the R/W/S/L part of the test</i>	Likert x 4			
			Test Value	<i>Studying for the TEAP Test helped me acquire skills that will be useful at university</i>	Likert
			Test Value	<i>I think that university English entrance exams should include a speaking test</i>	Likert
				<i>Why (not)?</i>	Open
Motivation	<i>I am motivated to study R/W/S/L for the test</i>	Likert x 4	Motivation	<i>The experience of taking the test motivated me to study R/W/S/L more in the future</i>	Likert x 4
Test Difficulty	<i>I think the R/W/S/L section of the test will be difficult</i>	Likert x 4	Test Difficulty	<i>The R/W/S/L section of the test was difficult</i>	Likert x 4

			Preparation sessions	The preparation sessions were useful for understanding the test	Likert
			Preparation sessions	As a result of the preparation sessions I was motivated to study R/W/L/S	Likert x 4
			Preparation sessions	Did you change the way you study English R/W/L/S as a result of the preparation sessions?	Likert x 4
				If so how?	Open x 4
			<b>Test preparation</b>		
			<b>In preparation for the test...</b>		
Vocabulary	I study academic vocabulary	Likert	Vocabulary	I studied academic vocabulary	Likert
Vocabulary	I recall the meaning of words so that I can understand them in reading and listening	Likert	Vocabulary	I recalled the meaning of words so that I could understand them in reading and listening	Likert
Vocabulary	I memorize new words and phrases so that I could use them in speaking and writing	Likert	Vocabulary	I memorized new words and phrases so that I could use them in speaking and writing	Likert
Grammar	I review/learn new grammatical forms in English for use in reading and listening	Likert	Grammar	I reviewed/learned new grammatical forms in English for use in reading and listening	Likert
Grammar	I review/learn new grammatical forms in English for use in speaking and writing	Likert	Grammar	I reviewed/learned new grammatical forms in English for use in speaking and writing	Likert
			<b>In preparation for the reading part of the test...</b>		
Reading	I read carefully to understand texts	Likert	Reading	I read carefully to understand texts	Likert
Reading	I skim read texts for the main ideas	Likert	Reading	I skim read texts for the main ideas	Likert
Reading	I search texts quickly for specific information	Likert	Reading	I searched texts quickly for specific information	Likert
Reading	I read notices, advertisements, emails and blogs in English	Likert	Reading	I read notices, advertisements, emails and blogs in English	Likert
Reading	I read academic texts in English	Likert	Reading	I read academic texts in English	Likert
Reading	I read educational materials and documents	Likert	Reading	I read educational materials and documents	Likert
Reading	I read and interpret graphs and charts annotated in English	Likert	Reading	I read and interpret graphs and charts annotated in English	Likert
Reading	I read stories and novels in English	Likert	Reading	I read stories and novels in English	Likert
Reading	I read about universities, university programs, and campus life online in English	Likert	Reading	I read about universities, university programs, and campus life online in English	Likert
Reading	I try to read quickly, under time pressure	Likert	Reading	I tried to read quickly, under time pressure	Likert
Reading	I analyze my answers to reading comprehension questions and tried to understand why answers were incorrect	Likert	Reading	I analyzed my answers to reading comprehension questions and tried to understand why answers were incorrect	Likert
Reading	I try to read English at least a little every day	Likert	Reading	I tried to read English at least a little every day	Likert
			<b>In preparation for the listening part of the test...</b>		
Listening	I take notes while listening	Likert	Listening	I took notes while listening	Likert
Listening	I listen carefully to catch the details	Likert	Listening	I listened carefully to catch the details	Likert
Listening	I listen to understand the main ideas	Likert	Listening	I listened to understand the main ideas	Likert
Listening	I listen to academic lectures in English on the Internet	Likert	Listening	I listened to academic lectures in English on the Internet	Likert
Listening	I listen to descriptions of charts and graphs	Likert	Listening	I listened to descriptions of charts and graphs	Likert
Listening	I listen to dialogues between two people	Likert	Listening	I listened to dialogues in English	Likert
Listening	I listen to dialogues between three or more people	Likert	Listening	I listened to dialogues between three or more people	Likert
Listening	I use the Internet to access audio/video in English	Likert	Listening	I used the Internet to access audio/video in English	Likert
Listening	I watch TV programs, TV dramas, movies and so on in English	Likert	Listening	I watched TV programs, TV dramas, movies and so on in English	Likert

Listening	<i>I listen to online audio and video material about universities and university courses in English</i>	Likert	Listening	<i>I listened to online audio and video material about universities and university courses in English</i>	Likert
Listening	<i>I try to listen to English at least a little every day</i>	Likert	Listening	<i>I tried to listen to English at least a little every day</i>	Likert
	<b>When I study English...</b>			<b>In preparation for the speaking part of the test...</b>	
Speaking	<i>I speak with little preparation time about personal topics</i>	Likert	Speaking	<i>I spoke with little preparation time about personal topics</i>	Likert
Speaking	<i>I speak with little preparation time about abstract topics</i>	Likert	Speaking	<i>I spoke with little preparation time about abstract topics</i>	Likert
Speaking	<i>I respond to personal questions</i>	Likert	Speaking	<i>I responded to personal questions</i>	Likert
Speaking	<i>I ask questions about various topics</i>	Likert	Speaking	<i>I asked questions about various topics</i>	Likert
Speaking	<i>I give my opinion about various issues and give reasons and examples</i>	Likert	Speaking	<i>I gave my opinion about various issues and give reasons and examples</i>	Likert
Speaking	<i>I role-play interviews</i>	Likert	Speaking	<i>I role-played an interview</i>	Likert
Speaking	<i>I practice speaking about a topic without stopping for a minute or more</i>	Likert	Speaking	<i>I practiced speaking about a topic without stopping for a minute or more</i>	Likert
Speaking	<i>I record my speaking and review it to try to improve</i>	Likert	Speaking	<i>I recorded my speaking and reviewed it to try to improve</i>	Likert
Speaking	<i>I use new words and phrases that I have learned when I speak</i>	Likert	Speaking	<i>I used new words and phrases that I have learned when I speak</i>	Likert
Speaking	<i>I mainly speak by myself aloud / by myself in my head / with a partner</i>	Likert	Speaking	<i>I mainly spoke by myself aloud / by myself in my head / with a partner</i>	Likert
Speaking	<i>I pay a lot of attention to speaking correctly</i>	Likert	Speaking	<i>I paid a lot of attention to speaking correctly</i>	Likert
Speaking	<i>I try to use complex grammar and more specific (higher level) words/phrases when I speak</i>	Likert	Speaking	<i>I tried to use complex grammar and more specific (higher level) words/phrases when I speak</i>	Likert
Pronunciation	<i>I try to improve my pronunciation</i>	Likert	Pronunciation	<i>I tried to improve my pronunciation</i>	Likert
Pronunciation	<i>I try to improve my intonation of sentences and questions when speaking</i>	Likert	Pronunciation	<i>I tried to improve my intonation of sentences and questions when speaking</i>	Likert
Speaking	<i>I try to speak in English at least a little every day</i>	Likert	Speaking	<i>I tried to speak in English at least a little every day</i>	Likert
	<b>When I study English...</b>			<b>In preparation for the writing part of the test...</b>	
Writing	<i>I write longer texts (200 words or more)</i>	Likert	Writing	<i>I wrote longer texts (200 words or more)</i>	Likert
Writing	<i>I write summaries of texts</i>	Likert	Writing	<i>I wrote summaries of texts</i>	Likert
Writing	<i>I write summaries of multiple texts and/or charts</i>	Likert	Writing	<i>I wrote summaries of multiple texts and/or charts</i>	Likert
Writing	<i>I write summaries of visual information (e.g., graphs, charts)</i>	Likert	Writing	<i>I wrote summaries of visual information (e.g., graphs, charts)</i>	Likert
Writing	<i>I review my own writing and tried to find things to improve</i>	Likert	Writing	<i>I reviewed my own writing and tried to find things to improve</i>	Likert
Writing	<i>I rewrite my writing to correct mistakes and improve it</i>	Likert	Writing	<i>I rewrote my writing to correct mistakes and improve it</i>	Likert
Writing	<i>I try to use new words and phrases that I have learned</i>	Likert	Writing	<i>I tried to use new words and phrases that I have learned</i>	Likert
Writing	<i>I pay a lot of attention to the content / organization of my writing</i>	Likert	Writing	<i>I paid a lot of attention to the content / organization of my writing</i>	Likert
Writing	<i>I try to write under time pressure</i>	Likert	Writing	<i>I tried to write under time pressure</i>	Likert
Writing	<i>I try to use complex grammar and more specific (higher level) words/phrases when I write</i>	Likert	Writing	<i>I tried to use complex grammar and more specific (higher level) words/phrases when I write</i>	Likert
Writing	<i>I pay a lot of attention to writing accurately</i>	Likert	Writing	<i>I paid a lot of attention to writing accurately</i>	Likert
Writing	<i>I try to write in English at least a little every day</i>	Likert	Writing	<i>I tried to write in English at least a little every day</i>	Likert
Other	Please tell us if there was anything else that you did a lot of in preparation for the test, which is not mentioned above.	Open	Other	Please tell us if there was anything else that you did a lot of in preparation for the test, which is not mentioned above.	Open
				<b>In preparation for the test...</b>	



Socio-Affective	<i>I sought advice about the test from a teacher or other</i>	Likert x 2
Socio-Affective	<i>I received feedback on my speaking from a teacher or others</i>	Likert
Socio-Affective	<i>I received feedback on my writing from a teacher or others</i>	Likert
Socio-Affective	<i>I spoke with other people in English</i>	Likert
Socio-Affective	<i>I worked with a peer to review each other's writing</i>	Likert
<b>In preparation for the test...</b>		
Test Strategies	<i>I did (no / 1-3 / 3+) practice tests in preparation</i>	Yes, No
Test Strategies	<i>Which of the following did you use in preparation?</i>	TEAP-related books; TEAP online material; Center Test materials
Test Strategies	<i>Anything else?</i>	
Test Strategies	<i>I spent time analyzing the test</i>	Likert
Test Strategies	<i>I tried to identify frequent question types</i>	Likert
Instruction	<i>I did not practice R/W/S/L for the test because I studied it a lot in classes</i>	Likert x 4

## Appendix 2: Sample interview questions for interviews

### Interview 1: Current English learning experience, example questions:

- How do you typically study English in class?
  - What percentage of your study time do you spend on studying speaking, reading, writing, and listening?
  - How do you study reading, listening, writing and speaking?
    - What materials and/or tools do you use?
    - Do you usually work alone, in pairs or groups?
  - *Further questions from Green (2014): What are the 'Reading' activities that receive so much student attention? Why are these activities selected in regular English classes and in test preparation classes? Why is so little attention given to speaking (and other skills)?*
- How do you typically study English outside of class?
  - What percentage of your study time do you spend on studying speaking, reading, writing, and listening?
  - How do you study reading, listening, writing and speaking?
    - What materials and/or tools do you use?
    - Do you study alone or do you study/practice English with others?
- How do you typically study for English tests now?
  - Do you spend more time on certain English skills than others? Why or why not?
  - Have you ever taken an English test that is not a school test? If so, how did you study for that test?
    - What materials and/or tools do you usually use?
    - Do you study alone or do you study/practice English with others?

**Interview 2: Preparation done TEAP, reflection on the TEAP test, example questions:**

- How did you study for the TEAP test?
  - What *percentage* of your study time did you spend on studying speaking, reading, writing, and listening?
  - Why did (or didn't) you focus on each skill?
- How did you study reading, writing, speaking, listening for the test?
  - What *specific activities* did you do to prepare for the test?
    - What materials and/or tools did you use?
    - Did you study/practice by yourself or with others?
    - Did you practice speaking spontaneously by yourself or with someone for the test?
- How did you study grammar, vocabulary and pronunciation for the test?
  - What *specific activities* did you do?
    - What materials and/or tools did you use?

**Interview 3 (OR Focus group): Reflections on the TEAP test and influence on study for entrance exams/future studies in university, example questions:**

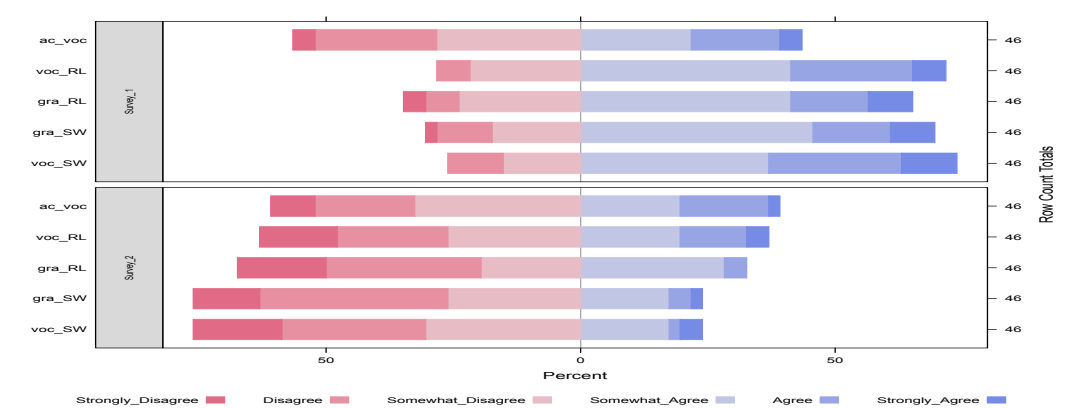
- How do you study English now?
- Did studying for the TEAP test influence how you studied for other entrance exams? (How?)
- Did studying for the TEAP test help you improve your English language skills?
  - How? (or why not?)
- Do you think entrance exams should have an English speaking test?
- Do you think the TEAP Test, or a similar four-skills test, should be used as the English entrance exam across universities in Japan?



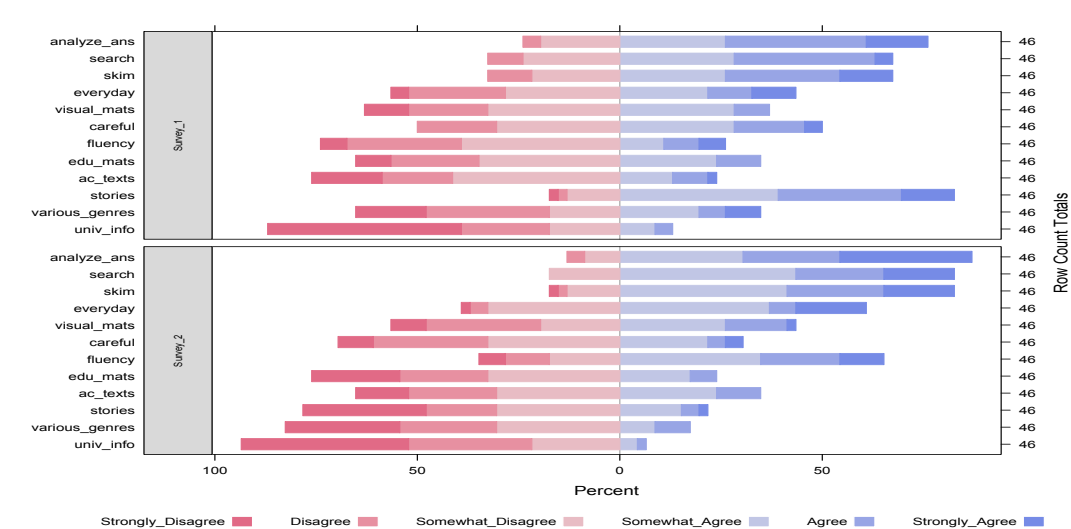
### Appendix 3: Open-ended responses for the question: ‘Did you change the way you study English R/W/L/S as a result of the preparation sessions?’

スピーキング
<p>友達や塾の大学生と英語を使って話してみた 録音するようになった。 違う言い方を考えるようになりました 教科書の音読をした 時間内に話すようにした。 様々な意見を持つこと 相手の言っていることの聞き取り方 自分の意見をどうしたら相手に納得させるかを考えるようになった。 過去問を解いた。 面接官に聞かれたと言う設定で話す練習を取り入れた CDのような流暢な音読を心がけるようになった。 回答を考える際二つ以上の理由をつけるように気を付けた。</p>
ライティング
<p>文法復習をした 書いた後の採点復習を大切にするようになりました。 学習を行う頻度が増えた 意見を述べるというよりも、文章の要約をする勉強を意識した。 型を意識すること 英作文対策もやるようになった。 学生生活で使うような単語を確認した。 構成(譲歩を取り入れる) 要点をまとめて英語で書く勉強に変わった 定期的を書くようになった 過去問を解いた。 全体の構成と意見の一貫性を考えるようになった。 そこまで変わらなかったが、学校の授業でやっていたように文と文の接続を考え、主張が一貫した客観的な文章が書けるよう気を付けた。</p>
リーディング
<p>洋書や長文を少し多く読むようにした 語彙を増やした。 解いた文章を音読するようにし、しっかり理解できるようにしました とにかく単語を頭に入れた 以前より多くの文章を読む機会を作った 合格した準一級の勉強を続けるようにした。 選択肢を先に読むこと 速読で趣旨をつかめる練習 アカデミックワードを覚えるようにした 過去問を解いた。 語彙を大切にするようにした 必要な情報を意識して読むようになった。 パラグラフの大意を掴む練習をした。 読み方を意識するようになった</p>
リスニング
<p>(英語の)CDを聞く時間を増やした メモの取り方を変えた。 音声台本を見ながらも一度聞いたら、音に合わせて読んで見たりするようになりました 設問の取りかかり方 教科書の音声聞く頻度を増やした 日常生活というより、学校でのリスニングに対応できるよう「TEAP 実践問題集」(旺文社刊)を何度も繰り返して聴いた。 英語のラジオなどを聞くこと リスニング問題の目の通し方 過去問を解いた。 数値に注意するようになった。 頻度を増やした 色々聞くようになった</p>

**Appendix 4:** Additional Figures for sub-skills survey data



*Figure 4: Focus on vocabulary and grammar in general (Survey 1) and in preparation for TEAP (Survey 2)*



*Figure 5: Focus on reading skills in general (Survey 1) and in preparation for TEAP (Survey 2)*

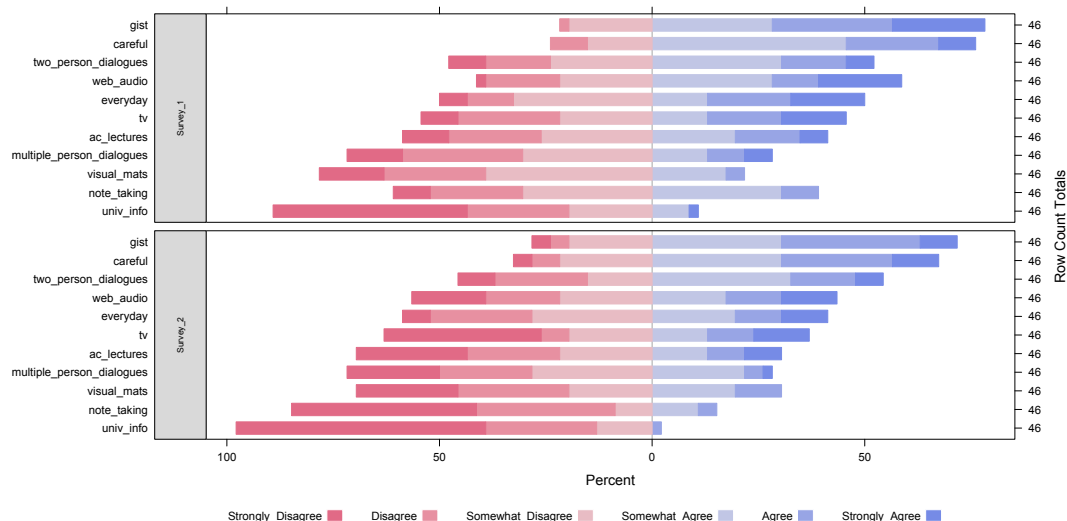


Figure 6: Focus on listening skills in general (Survey 1) and in preparation for TEAP (Survey 2)

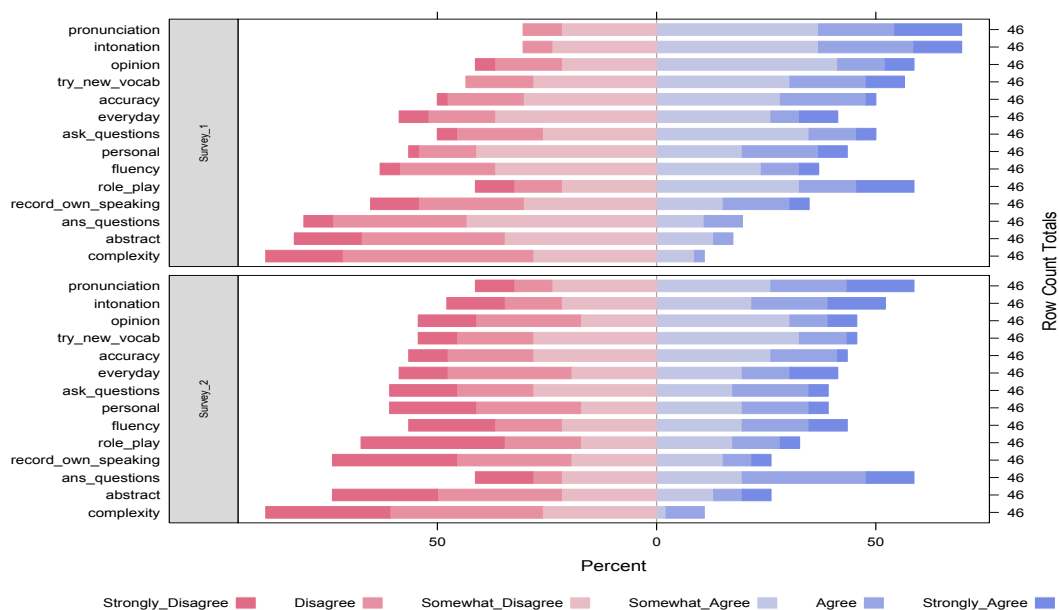


Figure 7: Focus on speaking skills in general (Survey 1) and in preparation for TEAP (Survey 2)

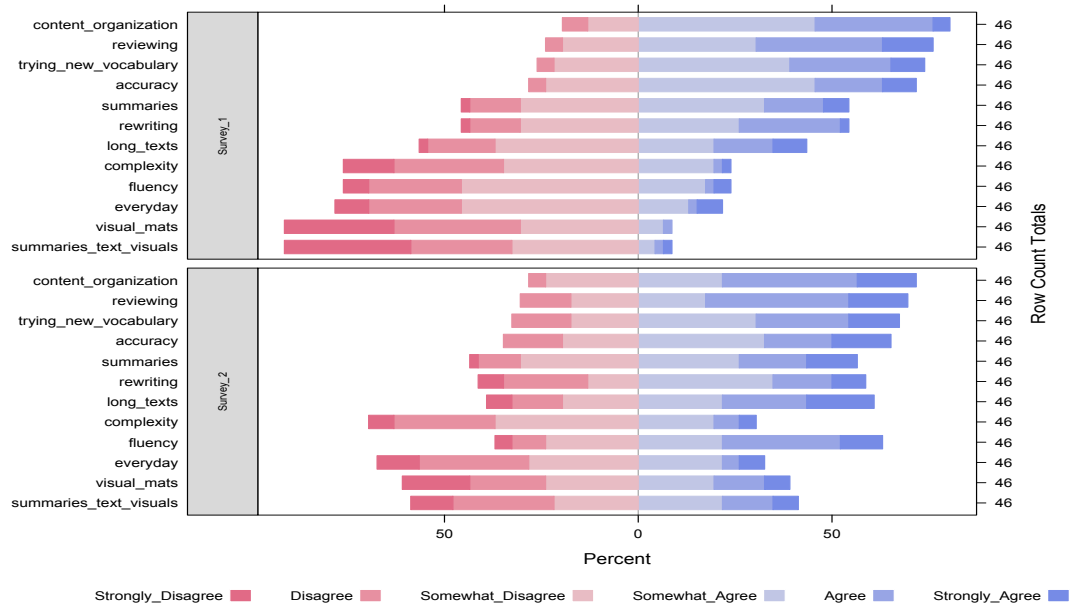


Figure 8: Focus on writing skills in general (survey 1) and in preparation for TEAP (survey 2)

**Appendix 5: Participants' agreement (1=strongly disagree, 6= strongly agree) and comments on the statement: 'I think that university English entrance exams should include a speaking test'**

**Agreement**

**(1-6) Open-ended responses to 'Why do you think so?' (38/46 participants responded)**

- 1 スピーキングでは自分の考えなどを述べさせられることが多く、解答がないので公正性に欠けると感じるから
- 1 現在の入試では全く使わないから
- 2 国立理系受験で使わないから
- 2 スピーキングの能力を入試ではかるのは難しいと思うから。
- 3 他の科目の勉強もあり、時間をあまり割けないから。
- 3 書き言葉と口から出る言葉は違う気がするから
- 3 四技能において最も実用的なものだと思うから
- 3 話せることこそ本当の英語力だと感じるから
- 4 将来的に英語を使う上でスピーキングはかなり重要だと感じるから。
- 4 英語力は読み書きだけではないと思うから。また、社会でよく使う技能はスピーキング力だと考えるから。
- 4 書類上で英語が使えても、話せないという人も多いから。
- 4 文法ができて話せないという意味が無いから
- 4 英語を用いてコミュニケーションをとるには話す技能も必要だと思うから。
- 4 英語はツールだから、習熟度を測るには「実際に使えるか」をその場で示せることが必要  
スピーキングはより実用的な英語力を測れると思うが、一方で対面の会話が苦手な人は英語力以外の面で躓いてしまう可能性があることを考えると会話試験中心のものは難しいかと思う。その大学がどのような傾向の生徒を取りたいのかに応じて他のスキルと合わせ、総合的に評価をつける必要があると考える。
- 4 英語をどんなに勉強しても話せなければコミュニケーションがとれないから。
- 5 海外で研究を進める時に役に立つから
- 5 文法が出来ても話せなかったら社会に出た時にあまり役に立たないと思うから
- 5 実際に「使う」ことに焦点を当てるべき
- 5 英語においてスピーキングは最も重要だと思うから
- 5 将来英語で話すことは必要だから。
- 5 コミュニケーションにおいて話す技能が最も求められると思うから。
- 5 英語は話せなければ意味が無いと思うから。
- 5 多くの交流は会話を通じて行われるものだから。
- 5 実際にコミュニケーションが取れることが大切だから。
- 5 今まで勉強してきた英語を、実践的に使いこなせるかを測るには一番効果的な方法だと思うから
- 5 会話できて初めて言語を習得したと言えると思うから
- 5 スピーキングは、将来就職した後でも役に立つスキルだから。
- 6 世界を舞台にしごとおしていく上では一番大事だから
- 6 英語は4技能使うことが必要不可欠になると思うから
- 6 現地では話せないとコミュニケーションが取れないので、何よりも大切なスキルだと思うから。
- 6 実際に使用する機会が多いのはスピーキングだから
- 6 英語の知識があっても大学生になったら話せることも大切だと考えるから
- 6 教科書だけの英語にとどまらず、実践的な英語を身に付けるため
- 6 英語のテストの点が良かったとしても、話せないと意味がないから。  
英語を読めて書けるのに話せないのではもったいないし、外国の人と話すことが必要な場面がこれからもずっと出てくると思うから。
- 6 話すことができれば相手の主張は理解できても自分の主張ができないため。
- 6 コミュニケーションは会話が基本だから。

