

Japanese cram schools and entrance exam washback

David Allen

Foreign Language Education Centre, Ochanomizu University

University English entrance exams are commonplace in Japan and competition for entrance to high-ranking institutions is fierce. Given the high-stakes nature of such exams, high school leavers typically attend exam preparation courses at *juku* cram schools. *Juku* are part of the shadow education sector, which has a huge presence in the East Asian region but which has escaped the attention of language education researchers. Moreover, while the entrance exams have been summarily criticized, limited empirical research exists into the washback effect upon students' language learning. The present study constitutes a medium-scale investigation into the learning behaviour, experiences and language proficiency of 133 undergraduate students at a prestigious Japanese university, all of whom had attended *juku* prior to entering university. Survey and interview data clearly show a focus on reading, grammar, and test-taking techniques, and a complete lack of focus on speaking skills. The focus on receptive abilities is also reflected in language proficiency scores on IELTS Tests taken during the first year at university. Washback from the entrance exams on language learning is discussed within the context of *juku* cram schools in Japan.

Keywords: Washback; Japan; shadow education; cram school; university entrance exam; IELTS

Introduction

Every year millions of students around the world sit high-stakes examinations that determine university admission. In Asia, the university entrance exam phenomenon occurs on an unprecedented scale. In the People's Republic of China, over nine million students annually sit the *gaokao* (National College Entrance Examination). College entrance exams are also firmly rooted in the educational cultures of Hong Kong, Taiwan, Korea and Japan. In Japan, high school leavers¹ go through *shiken jigoku* (examination hell) in preparation for entrance to higher education institutions. This period is seen as a crucial stage in determining an individual's future, and the act of cramming in preparation for these exams is widely accepted in these societies that are built upon Confucian beliefs in the value of education and diligence (Bray, 2009).

Although entrance exams and cramming are accepted as part of the social fabric of Japan (Locastro, 1990), they are not without their critics. From an educational standpoint, such high-stakes testing tends to encourage learning of short-term strategies, such as memorization of content, rather than developing a greater understanding of underlying principles (Green, 2013). Languages, notably English (Locastro, 1990), are core subjects in most university entrance exams, and have also been the focus of much criticism over the last two decades (e.g., Brown & Yamashita, 1995; Kikuchi, 2006). For instance, the lack of information about the design of the tests, such as test purpose and specifications, or research into the tests' validity, is also hugely problematic for their evaluation. These exams also vary greatly from university to university in terms of

their content and the type of tasks employed (Brown & Yamashita, 1995). Consequently, students hoping to enter higher-ranking universities need to be familiar with the particular format of the exam they will take, which creates an additional burden for students who apply for more than one university. In addition, due to the diversity and complexity of test items, the exams tend to promote test-taking techniques, or “testwiseness”. Finally, as commonly found in large-scale testing situations, the tests primarily assess receptive skills, especially reading, while assessment of productive skills is much more restricted, particularly speaking, which is rarely tested at all (Brown & Yamashita, 1995; Kikuchi, 2006).

Many of these critical issues centre on the notion of washback, in other words, the beneficial and/or detrimental effects upon teaching and learning that the test creates. It is somewhat surprising therefore that there has been little empirical research into washback in this context. The present study addresses this issue, by investigating washback from an established high-stakes university entrance examination on the learner and learning outcomes (i.e., language proficiency). Moreover, this study focuses specifically on learning at cram schools in Japan known as *juku*², which assist test-takers in their preparation for the entrance exams.

Juku and shadow education

The cram school industry is well represented throughout East Asia and also increasingly in other parts of the world (Bray, 2009). Such private tutoring has been termed *shadow education* (Bray, 2009) because it mimics mainstream education; as the mainstream system changes, so does the shadow system. In Japan, the Republic of Korea, and Taiwan, shadow education focuses primarily on teaching examination skills (Kim & Chang, 2010; Liu, 2012; Roesgaard, 2006), which may not be adequately covered in mainstream education. However, cram schools provide a range of educational services. In Japan, they provide tutoring for most subjects that appear on the entrance exams for major universities. They also prepare students for elite elementary, junior high and high school entrance exams, and they provide support for students who are falling behind at school (Roesgaard, 2006). The scope of the industry is phenomenal: Dierkes (2010) reports that there are over 50,000 cram schools in Japan and in 2010, Japanese households were spending approximately 924 billion yen (US\$12 billion) on private tutoring (Dawson, 2010, p. 16).

One aspect of shadow education that is commonly found in many countries is English Private Tutoring (EPT), which has begun to receive interest from researchers in TESOL in Jordan (Khuwaileh & Al-Shoumali, 2001), Germany (Mischo & Haag, 2002), Bangladesh (Hamid, Sussex, & Khan, 2009), the Republic of Korea (Lee, 2010) and Hong Kong (Yung, 2015). This research has focused on a range of topics such as experiences, attitudes, motivations, economic and social factors, achievements, and teaching and learning in EPT. Yung’s (2015) work is the most relevant to this study as he focused on the teaching and learning processes in EPT. In Hong Kong, he notes that “EPT courses are foregrounded as ‘cram courses’ aimed at developing examination competence” (p. 17) and not general English language proficiency. From interviews with students receiving EPT he found that teaching and learning followed an examination-oriented approach. That is, mock papers and model answers were standard materials, and teachers delivered lecture-style classes that were “loaded with content” (p. 10). Memorization of phrases for use in the exams and grammar drills were prominent learning activities and students perceived considerable development in their exam skills but were ambivalent about the usefulness of EPT for developing actual language proficiency. A limitation of Yung’s (2015) study, however, is that he did not

present information about the exams being studied for, such as the weighting of language skills on the tests, thus it is difficult to know whether the focus on non-communicative activities was driven by test content, or other factors. Moreover, the reported lack of development in communicative ability as a result of EPT cannot be verified due to the lack of objective measures of proficiency.

In Japan, given that *juku* have such a huge presence in English education, it is surprising that there has been so little research conducted into their practices. For instance, in a survey of issues of the Japanese Association of Language Teachers (JALT) Journal between 1979 and 2014, Lowe (2015) found only one article that focused on cram schools. Moreover, there seems to be no published studies that have investigated the relationship between the skills learned at *juku* and those tested on the entrance exams, which should be well aligned if *juku* also promote exam-oriented study.

Washback

Washback falls under the umbrella of test impact and consequences (Bachman & Palmer, 1996). Researchers focusing on washback have examined its intensity (Cheng, 2005) and the direction, in other words, whether the test leads to beneficial or damaging effects on teaching and learning, (i.e., positive or negative washback, respectively; Alderson & Wall, 1993). Hughes' (2003; also see Bailey, 1996) influential model of washback includes the test, participants, processes and products. Although there are a number of important participants, such as teachers, materials writers and parents, this paper focuses on the learner, who is arguably the most important participant in the process (Bailey, 1999). For the learner, the processes centre on the types of activities that learners do in preparation for the test and the type of skills and knowledge that are required when doing them. The product, or outcome, is learning, which can be measured using scores from the test in question or alternative measures of criterion-related abilities.

Research has shown that washback on teaching is not a simple linear process from test to teacher, but is mediated by a range of factors including teacher's beliefs, training and knowledge of the test (e.g., Watanabe, 2004). Although relatively few studies have investigated washback on learning (Cheng, 2005; Green, 2007, 2013), a growing number have revealed that this is also a complex process that is influenced by contextual and learner-related factors (Gosa, 2004; Tsagari, 2007). Washback to the learner (Bailey, 1999) can be assessed by investigating the extent of preparation, the skills and knowledge focused on, the learning activities, the motivation that students have as a result of the test, and the learning outcomes. The present study investigates these aspects of learning within the context of test preparation undertaken at *juku*.

Study context

This study focuses on the experiences of learners who studied for and successfully passed the entrance exam for the University of Tokyo (UT), which is Japan's most prestigious and competitive higher education institution. The UT English exam is a high-stakes test because the outcome determines access to higher education, and thus significantly affects test-takers' opportunities in life (Ishida, 1993).

An overview of the 2012 UT English exam, which was taken by these participants, is provided in the appendix. An analysis of the test questions suggests the approximate weightings for the 2012 UT English test were: Reading (70%), listening (20%), and writing (10%). There was no speaking component. Test items that focused on vocabulary and/or grammar, were usually embedded within a text and are thus counted

under the skill of reading. English to Japanese translation questions were also classified under reading because English language ability was tested in the form of comprehension.

Research questions

To investigate the aforementioned washback issues, three research questions were formulated:

Within the context of cram schools, to what extent do the university entrance exams influence:

1. the type of study that students engage in (i.e., focus on skills, knowledge, activities)?
2. learners' motivation to study particular skills?
3. learners' perceived and objectively measured language development?

Method

This study is part of a larger study (Allen, in press), in which Japanese undergraduate students' English learning and test preparation experiences, and language proficiency, were investigated.

Participants

Participants were recruited on a first come, first served basis via an open call to all first year students at the university. The larger study required all participants to take two IELTS tests and complete a survey immediately after the second test. Interviewees were recruited via the survey, selected randomly and interviewed within two weeks. The survey was completed by 190 second-year UT students, of whom 133 (70%) had been to cram school and whose data is thus used in this study. The mean age of the participants (89 male, 43 female, 1 no response) was 20 years. Nineteen of these participants were interviewed, of whom 13 (9 male, 4 female) had studied English at cram school during high school and were thus included in the present study. They were paid 1500-yen for participation.

Procedure

The survey was designed with reference to previous surveys and commentaries (e.g., Dörnyei & Taguchi, 2010), and administered in Japanese using online software (SurveyMonkey). All translations were double-checked by the author and research assistants. The final survey was revised following feedback from two learner focus groups (each with two to four participants) and two experts in the field. See Allen (in press) for more details of the survey design and administration.

Survey content

The survey contained questions about cram school study history, classroom organization, interaction between learner(s) and the teacher, learning activities, motivation and perceived development. Two categorical questions about cram school study history (Tables 1 and 2) were followed by fifteen likert scale questions about cram school study experience (Table 3).

Table 1. Responses to question:
How long did you attend juku for?

Period	% of responses
< 6 months	9
< 12 months	23
1-2 years	17
2+ years	51

Table 2. Responses to question:
*How many hours homework did
you study for juku?*

Hours	% of responses
<1	22
2	33
4	27
6	10
8	2
10	2
10+	4

Interview data

Two trained research assistants, who were postgraduate students at UT, conducted one-to-one, face-to-face semi-structured interviews. Each participant was interviewed once. Interviews were conducted and transcribed in the participants' first language, Japanese. Interview questions were in line with the survey items (Tables 1, 2 and 3) but were open-ended to elicit richer and more personalized responses (e.g., *Tell me about your English study experience at cram school. Which of the four skills did you focus most on at cram school? What types of activities did you do a lot of at cram school?*) Interviewers had the survey responses to hand to follow up as necessary.

Transcript data was organized manually in an Excel spreadsheet into the broad topics covered in the interviews. The spreadsheet data and the full transcripts were analysed reiteratively to identify recurrent and related information, and notes were taken about themes, including both confirmatory and dis-confirmatory information. Through discussion with interviewers and another researcher specializing in qualitative methods, adjustments were made and the themes within the data were verified. Both the interviewers and researcher are familiar with the Japanese pre-tertiary and tertiary educational contexts, which allowed them to take an insider's perspective. However, the researcher, while familiar with Japanese education, had been educated outside of Japan, and thus could identify themes that would have otherwise potentially been obscured by over-familiarity with the context.

Table 3. Survey responses (% of respondents)

		Strongly disagree	Disagree	Somewhat disagree	Somewhat agree	Agree	Strongly agree
My teachers talked for most of the class		8	11	9	12	17	43
I spent a lot of time working in:	Pairs	83	9	1	5	1	1
	Groups	77	5	6	6	2	4
	Individuals	2	1	5	5	13	74
My teacher(s) spoke mainly in English		63	20	9	2	3	3
I used English most of the time when speaking to the teacher		77	10	8	1	2	2
I used English most of the time when speaking to other students during classes		86	10	3	1	0	0
My homework often involved :	Reading	1	3	8	11	27	50
	Listening	23	14	23	17	12	11
	Writing	5	7	22	26	16	24
	Speaking	80	9	7	1	2	1
Class activities often focused on :	Reading	1	2	8	10	28	51
	Listening	16	12	19	19	19	15
	Writing	4	7	21	23	20	25
	Speaking	79	7	3	5	3	3
Class activities often focused on :	Pronunciation	41	20	15	11	9	4
	Grammar	4	4	4	17	26	45
	Vocabulary	2	3	10	18	31	36
I spent a lot of time doing :	Reading tasks	5	4	5	9	21	56
	Listening tasks (2 speakers)	18	16	23	23	9	11
	Listening tasks (3+ speakers)	25	16	20	22	9	8
	Writing about visual information	29	14	26	17	7	7
	Writing essays	14	13	21	24	14	14
	Speaking about everyday topics	79	7	8	4	2	0
	Speaking about abstract topics	83	7	5	3	1	1
I was often expected to speak in English immediately with little or no preparation time (fluency)		74	12	5	3	3	3
I studied test-taking techniques a lot		7	8	10	17	21	37
I was satisfied with the classes		9	6	13	29	21	22
I was motivated to study English while at cram school		7	6	10	19	29	29
I was motivated to study _____ as a result of the classes:							
	Reading	10	10	14	19	20	27
	Listening	18	10	19	19	15	19
	Writing	16	9	19	21	16	19
	Speaking	63	14	9	5	5	4
I think my proficiency increased a lot in:	Reading	2	5	6	18	26	43
	Listening	20	14	15	21	17	13
	Writing	7	14	19	22	18	20
	Speaking	72	12	6	7	2	1

Note: Mode responses are highlighted in bold.

Results

Survey results

Responses to the two categorical items (Tables 1 and 2) showed that 51% of participants spent more than two years, 40% between six months and two years, and 9% less than six months attending cram school. The majority (60%) of participants did between two and four hours of cram school homework per week, with 22% doing less than an hour and the remainder doing over six hours (18%). This amounts to considerable investment in cram school study.

Responses to the likert scale items (Table 3) indicated that students studied in classroom environments that were teacher-centred, had students working individually, and involved very little communication in English. Reading skills, grammar, vocabulary and test techniques were all highly prominent. In contrast, speaking skills and the related components, spoken fluency and pronunciation, received very little attention. Writing and listening skills and activities received more varied responses indicating some variation depending on the cram school.

Most participants were at least somewhat satisfied with their experience of cram school and were generally motivated to study English at the time. Learners were mainly motivated to study reading and least motivated to study speaking, with writing and listening receiving similarly varied responses. Perceived development in the four skills followed a similar pattern. Overall, the responses to the survey indicate that learning behaviour, motivation and perceived development regarding the four skills were generally in line with the weighting of the skills on the exam.

Interview data

Most interviewees attended cram school during their third year of high school, though some started earlier or continued for a year following high school graduation. The cram schools attended were all typical large-group, lecture style classes, except one (P11)³, which was in the form of a correspondence course. Nine participants attended *juku* that focused specifically on the UT exam and which were characterized by their almost exclusive use of UT past exams. A typical lesson followed this sequence:

- students did part of the exam within the time limit
- students submitted their answers
- all answers were then explained by the teacher
- scores (and sometimes feedback) were provided the following week.

Students rarely worked in groups, but if they did, the discussion was in Japanese. The interviews clearly indicate that preparation for the English entrance exam was students' sole objective, and there was little or no mention of study behaviour geared towards gaining communicative competence in English or in promoting language learning that was not directly relevant to the test, for example:

P1: Mainly we used past UT exam questions, or ones that were similar, and did those every week with the aim of improving our scores.

All 13 interviewees reported focusing on reading mainly. The amount of reading material, and the speed at which they were required to read, was mentioned regularly and two students reflected that this helped to improve their comprehension of English texts.

Students tended to reflect on the significant attention paid to English grammar, generally through mechanical multiple-choice test items accompanied by teacher-centred instruction. One student's description illustrates this and shows how teachers focused on grammatical aspects that could potentially confuse learners, in other words, things that may cause problems in the exam:

P5: Firstly, for grammar, each person did 30 (multiple choice) questions really quickly, then the teacher went over important grammar points, like the difference between sensible and sensitive. He'd go through them all really quickly and everyone would be scribbling down notes, and we'd just keep going like that.

Interviewees' reports of their listening study varied a lot, in line with the survey responses, with some students studying at *juku*, but many doing test questions by themselves at home. Three students (P7, P8, P12) explained that as few entrance exams contain a listening component, it is rarely focused on at *juku*.

Students did a lot of writing practice, mainly at home and primarily writing responses to UT test questions, but sometimes using other universities' test questions. One noted (P8) that writing was focused on about 10-20% of the time, which is uncannily similar to the weighting of the writing component on the UT test. The teacher commonly analysed a model answer of a task (and sometimes student's answers) and students learned techniques to compose a similar responses. Whether or not students received feedback from teachers differed considerably from school to school. One student (P12) asked her high school teacher to provide feedback as she could not receive any at cram school, while others got it at cram school (P2-5, P7), some only if they asked for it (P8, P9). This shows a lack of priority placed on writing in line with its minor role in the test.

Speaking and its related aspects, such as pronunciation and spoken fluency, were hardly mentioned by the respondents. Two students (P6, P7) stated that the lack of focus on speaking was because it does not appear on the test. Only one interviewee (P5) studied at a school that specializes in English and holds oral communication classes with native speakers. However, these classes tended to involve pair dialogue drills in which accuracy appeared to be the focus rather than communication. As the excerpt below illustrates, her experience positions speaking as a minor aspect of the course, sandwiched between more important hours of answering multiple-choice questions.

P5: The style was like [...] just circling answers over and over, then we'd break and speak to a native speaker, we had just a little time to speak English, and then after that it was back to circling, circling, circling.

IELTS Test data

The third research question asked whether students' language proficiency after having entered UT reflects the pattern of study done in preparation for the entrance exam. To investigate this, the scores from an initial IELTS Test taken by all 133 learners within 11 months of entering the university were examined (Table 4). The mean scores on the four skills follows the same pattern as the focus of study as elicited through the survey and interviews. Learners scored almost two bands higher on reading compared to writing, and over one band higher on listening compared to speaking. This imbalance in skills is unusual compared to IELTS test-takers worldwide including those with Japanese as a first language (see Table 4) and thus appears to be the result of preparation for the UT exam.

Although learners reported studying writing but not speaking at cram school, the respective scores on the IELTS test were similar. This may be because the single writing task on the UT test is very short (50-60 words), whereas the two tasks in the IELTS Writing exam are much longer (150 and 250 words, respectively); moreover, the task types differ greatly, with the IELTS Writing tasks being considerably more challenging.

Table 4. Participants' IELTS scores from test taken within one year of leaving *juku*

Test Component	IELTS mean scores of study participants (n=133)	IELTS 2012 worldwide mean scores	IELTS 2012 worldwide mean scores (Japanese L1 speakers)
Reading	7.1 (0.9)	6.0	5.9
Listening	6.5 (1.0)	6.0	6.0
Writing	5.5 (0.6)	5.5	5.3
Speaking	5.4 (1.1)	5.9	5.6
Overall	6.2 (0.7)	5.9	5.8

Note: Standard deviations are in parentheses. The average of female and male candidates' IELTS data in 2012 was taken from <http://www.ielts.org/researchers/analysis-of-test-data/test-taker-performance-2012.aspx>

Discussion and implications

The purpose of this study was to investigate washback within the context of *juku* cram schools in Japan. The washback to the learner, in terms of learning behaviour, motivation and outcomes, was found to be intense.

In regard to the first research question, washback upon learning behaviour was clearly observed: The activities done in classes and for homework were largely derived from past UT exams, and the focus on skills reflected their relative weightings on the exam. Multiple-choice grammar and vocabulary exercises were a central focus, as they feature prominently on the exam. Unsurprisingly, these classes were also teacher-centred, largely non-interactive, and involved a lot of test practice combined with teacher explanation of how to answer test items. The *juku* environment thus encouraged a test-taking mentality, and did not promote skills necessary for communication. These findings resonate with those reported by Yung (2015), though additionally the current study shows clearly how *juku* curriculum content is determined by test content. The *juku* curricula were narrowed to focus only on language skills, knowledge and activities that feature on the actual tests being prepared for.

In regard to the second research question, learners appeared to be motivated to study the skills according to their weighting on the test. Learners were extremely goal-oriented, practical, and test focused; they did what they perceived as necessary to succeed on the test. In this sense, the test provided an immediate incentive and thus motivating function (Dörnyei, 1998, p. 121). The *juku* classes encouraged the idea that practicing test questions, and extensive grammar and vocabulary questions, were the best ways to improve scores on the test. In other words, the content of the test became the curriculum of the *juku* course, which mediated the students' motivation for language learning.

Regarding the third research question, participants' perceived development across the skills, as well as their IELTS Test performance, were roughly in line with their focus of preparation and motivation. Here, the survey, interview and test data, all provide evidence that the UT test creates a symbiotic form of washback upon behaviour, motivation and learning outcomes within the context of cram schools in Japan.

An important caveat regarding learning outcomes at cram school is that high school teaching in Japan is also primarily directed towards skills and knowledge found on the entrance exams, especially during the year preceding the exams (Allen, in press; Green, 2014). Japanese high school teachers believe the success of their students on the entrance exams is important for the success of their schools (Green, 2014), which is a key reason for focusing on skills that feature on the exams. The skew towards receptive skills observed here is therefore not purely a consequence of studying at cram school, but of English learning within both the mainstream and shadow education systems. This is an inherent problem with researching shadow systems. Because they mimic the mainstream system, it is often not possible to independently assess the effects of either system upon learning (Hamid et al., 2009).

There are, however, two reasons for supposing a unique contribution of EPT in Japan. Firstly, high school teaching does not appear to focus on specific exams as students apply to a variety of higher education institutions, all of which have different entrance exams. Secondly, the level of high school English was reported by interviewees in the larger study (Allen, in press) to be too low for the UT exam, which led them to do their *juku* homework during high school English classes. Therefore, where mainstream education fails to prepare its students sufficiently for entrance to higher education institutions, shadow education steps in to fill the gap. Yung (2015) suggests this is a key function of EPT in Hong Kong, and the present study supports this view for the Japanese context.

Importantly, EPT does not create positive or negative washback; the potential for both is in the design of the test (Green, 2007). Tests, therefore, must be critically evaluated using a validity framework such as the socio-cognitive framework (Weir, 2005), in which washback is considered an aspect of consequential validity. In this way, test designers can predict learner behaviour and outcomes, and manipulate the potential for positive washback. Test designers must also provide information to all stakeholders regarding the purpose of the exam and the constructs that it is designed to measure. Currently, entrance exams in Japan publicly release little, if any, information about the tests. This leaves learners and teachers guessing the purpose of test questions, which is likely to increase the risk of misinterpretation and thus limit any positive washback that the test could otherwise achieve. This is crucial in the EPT context, where the curriculum is based directly on the content of the exam.

However, even when tests are designed with positive washback in mind, the actual effect may not be as predicted. A number of studies investigating the introduction of high stakes English language tests in the East Asia region suggest that teachers and learners inevitably adopt short-term strategies for success on the test, without considering whether these strategies will benefit the learners' language ability (e.g., Andrews, Fullilove, & Wong, 2002; Qi, 2007). Given that the role of EPT is to improve students' exam scores, and often within limited time frames, it is foreseeable that short-term strategies will continue to be taught and learned to some extent regardless of the intentions of the test designers.

The key issue for future research is to investigate how shadow systems adapt to the introduction of new exams, particularly when the exams are designed to measure communicative ability across four skills. This research should observe learners in EPT

environments in order to evaluate more precisely the washback deriving from specific test tasks, and to evaluate it as positive or negative depending on whether studying for the test leads to the development of communicative or test-taking abilities, respectively. Moreover, studies investigating the views of teachers, parents and other stakeholders will be invaluable for gaining a better understanding of the perceived role of EPT in educational contexts.

Conclusions

In the East Asian context, where mainstream education systems are highly examination-oriented and exam scores are highly valued, shadow education is likely to retain an important role. As long as high-stakes tests determine entrance to higher education in this context, learners and their parents will turn to shadow systems to increase the chances of success on exams, and in turn, in society. Within the context of EPT in Japan, entrance exams create an intense washback effect on learning and *juku* intensify this by narrowing the curriculum to the skills, knowledge and activities that are featured on the test. Test designers must be aware of the role of EPT while striving to produce exams that increase the likelihood of positive washback on learning.

Notes

1. The Japanese formal school system is divided into elementary (6 years, ages 6 to 12), junior high (3 years, ages 13 to 15) and high school (3 years, ages 16 to 18). Higher education applicants must initially take the National Center for University Entrance Examinations (NCUEE) exam, which is a syllabus-based examination that takes place in the final year of high school. Depending on their results for this test, applicants may be eligible to apply for higher education institutions, which then administer their own internally constructed entrance examinations.
2. Cram schools in Japan include *juku* and *yobiko*, which are very similar types of educational institutions. However, while *juku* cater mainly to high school students who attend after school, *yobiko* are often also attended full-time by students who have finished high school and are preparing for entrance exams. In the present paper, *juku* (and cram school) refers to both *juku* and *yobiko*.
3. Interviewees were identified using numbers P1 to P13.

Acknowledgements

This study was funded by the British Council though the findings and opinions presented herein are those of the author and do not necessarily reflect the views of the British Council, its related bodies or its partners. I am grateful to Akiko Katayama, Sayaka Meguro and Kimie Yamamura for assistance with the interviews and translations. I am also grateful to Diane Hawley-Nagatomo, and two anonymous reviewers for comments on an earlier version of this article.

About the author

David Allen is an associate professor in the Foreign Language Education Centre of Ochanomizu University in Japan. His research interests are in the fields of psycholinguistics (specifically, Japanese-English bilingual lexical processing) and applied linguistics (language testing, peer-feedback in second language writing, corpus studies).

References

- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129.
- Allen, D. (in press). Investigating Japanese university students' English language proficiency: Predicting factors and pre-tertiary washback. *IELTS Research Reports*.
- Andrews, S., Fullilove, J., & Wong, Y. (2002). Targeting washback—A case-study. *System*, 30(2), 207-223.

- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257-279.
- Bailey, K. M. (1999). *Washback in language testing (TOEFL Monograph Series 15)*. Princeton, NJ: ETS. Retrieved from <https://www.ets.org/Media/Research/pdf/RM-99-04.pdf>
- Bray, M. (2009). *Confronting the shadow education system: What government policies for what private tutoring*. Paris, France: UNESCO International Institute for Educational Planning.
- 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 University Press.
- Dawson, W. (2010). Private tutoring and mass schooling in East Asia: Reflections of inequality in Japan, South Korea, and Cambodia. *Asia Pacific Education Review*, 11(1), 14-24.
- Dierkes, J. (2010). Teaching in the shadow: Operators of small shadow education institutions in Japan. *Asia Pacific Education Review*, 11(1), 25-35.
- Dörnyei, Z. (1998). Motivation in second and foreign language learning. *Language Teaching*, 31(3), 117-135.
- Dörnyei, Z., & Taguchi, T. (2010). *Questionnaires in second language research: Construction, administration, and processing*. New York: Routledge.
- Gosa, C. M. C. (2004). *Investigating washback: A case study using student diaries*. (Unpublished PhD thesis), Department of Linguistics and Modern English Language, Lancaster University, Lancaster, England.
- 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. (2013). Washback in language assessment. *International Journal of English Studies*, 13(2), 39-51.
- 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.
- Hamid, M. O., Sussex, R., & Khan, A. (2009). Private tutoring in English for secondary school students in Bangladesh. *TESOL Quarterly*, 43(2), 281-308.
- Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.
- Ishida, H. (1993). *Social mobility in contemporary Japan: Educational credentials, class and the labor market in a cross-national perspective*. Stanford, CA: Stanford University Press.
- Khuwaileh, A. A., & Al-Shoumali, A. (2001). Private tuition in English: The case of two universities in Jordan. *English Today*, 17(1), 31-35.
- Kikuchi, K. (2006). Revisiting English entrance examinations at Japanese universities after a decade. *JALT Journal*, 28(1), 77-96. Retrieved from <http://www.jalt-publications.org/archive/ji/2006a/perspectives.pdf>
- Kim, J.-H., & Chang, J. (2010). Do governmental regulations for cram schools decrease the number of hours students spend on private tutoring? *KEDI Journal of Educational Policy*, 7(1), 3-21.
- Lee, B. (2010). The pre-university English-educational background of college freshmen in a foreign language program: A tale of diverse private education and English proficiency. *Asia Pacific Education Review*, 11(1), 69-82.
- Liu, J. (2012). Does cram schooling matter? Who goes to cram schools? Evidence from Taiwan. *International Journal of Educational Development*, 32(1), 46-52.
- Locastro, V. (1990). The English in Japanese university entrance examinations: A sociocultural analysis. *World Englishes*, 9(3), 343-354.
- Lowe, R. J. (2015). Cram schools in Japan: The need for research. *The Language Teacher*, 39(1), 26-31.
- Mischo, C., & Haag, L. (2002). Expansion and effectiveness of private tutoring. *European Journal of Psychology of Education*, 17(3), 263-273.
- Qi, L. (2007). Is testing an efficient agent for pedagogical change? Examining the intended washback of the writing task in a high-stakes English test in China. *Assessment in Education: Principles, Policy & Practice*, 14(1), 51-74.
- Roesgaard, M. H. (2006). *Japanese education and the cram school business: Functions, challenges and perspectives of the juku*. Copenhagen: Nordic Institute of Asian Studies Press.
- Tsagari, D. (2007). *Investigating the washback effect of a high-stakes EFL exam in the Greek context: Participants' perceptions, material design and classroom applications*. (Unpublished PhD thesis), Lancaster University, UK.

- Watanabe, Y. (2004). Teacher factors mediating washback. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 129-146). Hillsdale, NJ: Lawrence Erlbaum.
- Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. London: Palgrave Macmillan.
- Yung, K. W. H. (2015). Learning English in the shadows: Understanding Chinese learners' experiences of private tutoring. *TESOL Quarterly*, 49(4), 707-732.

Appendix: Overview of the 2012 University of Tokyo English exam

Note: The full test, answers and analysis are available at: <http://kaisoku.kawai-juku.ac.jp/nyushi/honshi/12/t01.html>. The information on the website and much of the test is written in Japanese.

Q	Skill	Task type	Receptive / Productive	Text length (words)
1A	Translation	Translate English text (256 words) to Japanese (70-80 characters) Add sentence to text (MC)	Receptive (Produce in Japanese)	256 313 (+ 363 for paragraph selection task)
1B	Reading	Add word to text (MC) Add paragraphs (MC) Choose best summary (MC) Sentence gap-fill items, first letter and definition/synonym provided (5 items)	Receptive	
2A	Reading (Vocabulary)	Write a 50-60 word response to the prompt 'If you could read people's feelings, what would be the result?'	Receptive / Productive	NA
2B	Writing	Picture completion (add word; 3 items)	Productive	NA
3A	Listening	Picture completion (add number) Select answer to questions (MC)	Receptive / Productive Receptive	599
3B	Listening	Select answer to question (5 items) (MC)	Receptive	504
3C	Listening	Select answer to question (5 items) (MC)	Receptive	614
4A	Reading (Grammar)	Word deletion from sentences (5 items)	Receptive	173
4B	Translation	Translate sentences from a text (3 items)	Receptive (Produce in Japanese)	259
5	Reading, (Grammar, Vocabulary)	Select synonym, add word, add sentence part, select best meaning (all MC; 8 items) Translate sentence (1 item)	Receptive	808

Note: MC = multiple choice