

---

# WASHBACK AND THE ASSESSMENT PRACTICES OF ESL INSTRUCTORS AT JAPANESE UNIVERSITY

**Lee Hughes**

Asia University, Tokyo, Japan

E-mail: leeh@asia-u.ac.jp

Received: 2021-01-18

Accepted: 2021-03-09

## Abstract

This study investigated awareness among native-English instructors teaching English to first year freshman students studying English at a university in Japan, and the potential effects of their chosen methods of oral evaluation on washback. Washback effect refers to the impact of testing – whether positive or negative - on teaching practices, curriculum design, and learning behaviors. Oral washback, in particular, in evaluation has seen increased importance in recent years, with more consideration given to how to provide positive oral washback in evaluations to ensure that they are as effective as possible in improving English levels. A mixed methods approach was used – a survey and interviews – with five instructors to find out about how they assess their students and how aware and how much attention they gave to possible washback effects of their chosen methods of evaluation. It was found that there is quite a variance among different instructors and knowing more about how to promote more positive oral washback to be useful in improving evaluation methods used.

**Keywords:** *ESL, washback, assessment, teacher-based, evaluation*

## 1. Introduction

Washback is primarily associated with assessment, while washback is usually referred to as the effect of a test on teaching and learning the language (Green, 2013; Saif, 2006; Saville, 2009; Takagi, 2010; Tsagari, 2012). More and more research is beginning to be conducted into the nature of washback, and it is already accepted that it refers to a range of unplanned and often complex phenomena that impact upon learning and teaching while occurring around assessment processes.

The topic of washback itself and how it applies to oral assessment procedures/ contexts is a relevant and important topic to the art of English language teaching, and there have been a number of studies on oral washback in evaluation of students (Khan & Stapa, 2019). This has led to increased understanding of how to promote more positive kinds of washback. Before, ESL/EFL instructors' classroom assessment practices were investigated to find out their purposes and methods (Cheng, Rogers, & Hu, 2004), one of particular interests investigating the washback of an oral assessment system in an EFL classroom (Muñoz, & Álvarez, 2010). In this article it was found, using a mixed methods approach, that positive washback was found in some areas, but that constant guidance and support is required to help teachers create positive washback in other areas.

Various studies into high-stakes tests have already revealed that they can have a very significant effect on teaching and learning in terms of languages being taught (Alderson & Wall, 1993; Green, 2013; Xie, 2013), and this can result in serious consequences for the students involved. There are, however, varying opinions among researchers in relation to the nature of washback. There are several (Alderson & Wall, 1993; Dawadi, 2018) who propose that language tests have direct effects on students' learning practices. This relationship may not always be direct between the relationship between a test and learning practices, and this may be so as language assessment is a social activity connected to a large number of variables in society which constantly overlap and interreact. As a result, we need to remember that language assessment is inevitably affected by the socio-cultural practices in a specific society in which the learning is taking place.

The studies and research that have already been conducted in relation to washback in assessment are vital to help ensure that we understand how aware teachers and instructors are as to the importance of promoting positive washback in relation to assessments they have their students do. By researching more about instructors' feelings and beliefs about assessment, and how they actually choose to conduct their assessments for their students – assuming they are able to do so – we can find out if more needs to be done, for example, in terms of improving teacher training, to ensure that assessments are used more positively in relation to learning. This research that was completed at a university in Japan attempts to add more data to this pertinent issue in ESL by collecting data in the form of a survey and also interviews with ESL instructors.

## 2. Literature Review

It is actually a relatively new idea that tests can have effects in evaluation and assessment in general, especially in relation to individuals. Only in the 1990s did the idea start to be researched significantly and understood better (Loumbourdi, 2014). The very usage of the term 'washback' (sometimes referred to as "backwash": Hughes, 2003; Prodromou, 1995) was only discussed widely around this time (Alderson & Wall, 1993), but the researchers acknowledged that it was still a very little understood concept that had not been properly explored and understood. It was not until Alderson & Wall's article (1993) that researches began to take the concept of washback seriously and begin to discover more about it.

Nowadays, there is a large amount of research that not only confirms the existence of washback, but also shows that it is a complex and multi-faceted phenomenon. Still, since there is not always a linear association between assessments and teaching/learning processes, it is not easy to say exactly what washback looks like in each situation (Dawadi, 2019).

The relationship between assessment, feedback, and learning has also been studied at the high school level (Davison, & Leung, 2009), in which issues in teacher-based assessment (TBA) were examined. While TBA assessment was largely considered to be positive, the article made clear that more thinking is needed around fairness, trustworthiness, and ethics when orally assessing students. More research into teacher training and professional development was also advised. Then it was concluded that TBA should be applauded for its strong potential to improve learning and facilitate teaching. This paper adds more to our understanding of where we are in terms of how some instructors feel about these issues and look at possible ways to increase positive washback and reduce negative washback.

Another interesting article investigated native and non-native teachers' judgments of oral English, in which Kim (2009) found through a mixed methods study that the judgments of native-speaker teachers were more elaborate and detailed than those of the non-native teachers in terms of pronunciation, specific grammar use, and the accuracy of transferred information. These findings could have implications in classes where non-native speakers are teaching the same students as native-speakers if oral English is being assessed. With training and efforts to improve washback, this may be possible to change.

Although focusing on writing, Matsuno (2009) also made some interesting observations in Japanese university EFL classrooms. This study found self-assessment was idiosyncratic and therefore of limited value as part of a formal assessment, but that peer-assessors were internally consistent and their rating patterns were not dependent on their own writing performance. This could also have implications for when these methods are used during oral assessments. Little evidence of bias was also found, and so peer-assessments may play a useful role.

Wicking (2017) conducted a similar study to this when he also researched about the assessment techniques of English teachers in Japanese universities. He pointed out that teachers are usually in the highest position of power regarding their ability to control how tests are created, conducted and interpreted. As a result of this, it is vital to understand about teachers' beliefs and practices regarding the very important issues around assessment. He discovered some differences in beliefs and practices of EFL teachers at Japanese universities, although for the most part he believed they were only to be considered slight. However, he concluded that educational leaders and policy-makers should use information like this to ensure better testing practice in the future.

In his study, he also noted that one reason for assessing students given by teachers was administrative: to determine final grades. While this is, of course, very important, it remains to be seen how much attention teachers are paying towards ensuring their testing procedures promote more positive than negative washback, and therefore play a useful part in the learning process. In Japan, in particular, it has been observed that the over-importance of testing can cause a number of negative effects (Sugimoto, 2014), but in his study most teachers felt that assessment contributed positively to EFL learning. However, part of the reasons for this was that they felt they had control over how to administer and conduct tests, and would have a much more negative opinion if they instead had to give standardized and high-stakes tests.

An interesting study was conducted on the washback effect and its relationship with fear of negative evaluation (Azmi, & Sham, 2018) in Malaysia. This was a mixed-method study not too much dissimilar from my own, and it showed that the students had a low fear of negative evaluation, and the positive washback was high in relation to the oral evaluation conducted. This shows that it is important to develop a stronger curriculum based around oral English language education, improving the training and skills for the English teachers.

Another Malaysian article explored the washback effects of the Malaysian University English Test (MUET) on teachers' perceptions on their classroom teaching and teaching materials. This study is in line with Cheng (1997) that high-stakes tests have both negative and positive washback effects on teaching, material used, and the depth and degree of teaching. It was found that the test influenced teaching and also what teachers taught, along with the degree and depth of their teaching. However, the study was not able to calculate the strength and weakness of the positive and negative washback effects presented by the test.

With regards to the IELTS examination, Estaji and Ghiasvand (2019) pointed out in their study of Iranian instructors, that washback had not yet been well researched in relation to high-stakes tests and teachers' professional identity. This was also a mixed method study, and they found that there was no improvement in IELTS instructors' professional identity total score from pretest to posttest, but there was a positive relationship between their past IELTS related experiences and their professional identity. The instructors also felt that they had no choice but to work for the test instead of promoting real literacy skills, and that to merely survive they had no option but to use past papers, test-taking strategies, and tips and tricks.

In Saudi Arabia, research on the washback effect of LOBELA (Learning Outcome Based English Language Assessment) was conducted at the first year of university students (Hazaca, & Tayeb, 2018). Again, the method used was a mixed method approach of a questionnaire and interview. In the study, they found that LOBELA washback had the greatest effect on teaching methods, and an equal effect on teachers' motivation and attitudes. At the end they concluded that there needed to be more training to develop language assessment literacy among EFL lecturers.

### **3. Research Method**

I have explored how to conduct an effective qualitative and quantitative mixed method survey with regards to oral assessment washback. A quantitative survey (as used by Cheng, Rogers, & Hu, 2004) was given to university instructors at the beginning of the semester to find out more about the instructors and their beliefs regarding assessment. The first section consists of their personal background, and the next section asks the instructors to rate their purposes and reasons for assessing their students. This includes asking what is their primary source for test items and assessment procedures, and how they provide feedback, if indeed they do so.

The quantitative survey used by Cheng, Rogers, & Hu (2004) was used here because in their paper they were able to demonstrate, in their finds, the complex and multifaceted points that are involved in different learning and teaching environments with regards to assessment. Through the surveys they were also able to obtain insights about the nature of assessment practices at the tertiary level, which is relevant to what I am trying to discover more about.

Initial questions included asking their age, gender, qualifications and experience, before moving on to ask how much they agreed or disagreed with a variety of purposes and reasons for assessing their students. Instructors were then asked to outline the methods they used to assess students' oral skills, such as oral reading, oral interviews, and oral presentations. The next section asked the instructors to identify their primary source(s) for test items and other assessment procedures, while the last part asked them to check the ways in which they provided feedback to their students during the course.

This provided useful information about the purposes and reasons why the instructors were assessing their students, their methods for checking their students' oral skills, and their procedures of assessment and evaluation. Using this information, it would be possible to understand the kinds of washback that might be expected in each case. However, to gain more information of each instructor's knowledge and beliefs, an interview was also conducted and recorded with each instructor at the beginning of the semester. Follow-up questions were asked occasionally if I felt it would lead to more interesting information being unearthed.

The interviews were carried out in compliance with the ethical standards of the British Educational Research Association (BERA, 2018). Both participants and their parents/caregivers gave their informed consent before any primary data was collected for this study. Each participant was told that participation was entirely voluntary and that all data would be kept private and anonymized as soon as possible. All participants were given a detailed account of the project in order to gain informed consent.

#### Interview

As for the qualitative interviews, I devised the following questions to gather more specific and detailed evidence of the instructors' feelings and thoughts regarding evaluation methods in general, and washback in evaluation:

#### Initial Questions

1. When I say the word assessment, what images or feelings does the word conjure?
2. How do you feel the assessment of your students' going?
3. What areas of English do you feel the most comfortable assessing?
4. Think about the last assessment you did. How do you think it went?

#### Planning Stage

5. How did you come about choosing this kind of assessment for this class?
6. How much freedom did you have in choosing this kind of assessment?
7. Who decides on which assessments you are able to use?
8. How did you know that it would be suitable for your students?
9. How do you integrate assessment with teaching and learning?
10. How did you prepare your students for the assessment?
11. How much freedom did the students have in how they did/approached the assessment?
12. How did you go creating the standards and criteria for the assessment?

#### Implementation State

13. Tell me how you introduced the assessment to the students.
14. How do you feel the going of the assessment?
15. Did you feel that you were giving feedback to the students during the assessment?
16. How did you feel on the students' work on the assessment?
17. Would you change anything if you were to do it again?

#### Monitoring Stage

18. How did you rate the assessment?
19. What will you do with the results?
20. Will you provide feedback? If so, what kind of feedback?
21. Who will have the access to the results of the test?
22. What effects did the assessment have on your teaching this semester?
23. How do you think that the result of the assessment will affect the students?

#### Dissemination Stage

24. How are these assessments compared to the others that you have done?
25. How have these students performed in relation to the others in the class?
26. Do you feel that the assessment was effective in judging the achievement of the students?
27. Are there any restrictions on how many students can achieve certain grades? Do you have any freedom in this regard?

#### 4. Results and Discussion

Three male instructors and two female instructors were given the questionnaire and took part in an interview. All of them were between the ages of 30-50, with Master's degree related to ESL/EFL/TEFL or linguistics, and had been involved in ESL/EFL teaching for at least ten years. In terms of the purposes and reasons for assessing their students, the points are to obtain information on students' progress, make students work harder, and determine the final grades and this is a part of the questionnaire. This information is to provide insight into the quantitative survey results and to be aware of the backgrounds of all five instructors collectively. All five interviews were transcribed, and pseudonyms have been utilized for direct quotations.

In terms of the methods for assessing their students' oral skills, oral presentations were conducted by four of five of the instructors, and three out of five did oral interviews and/or dialogues. Oral reading (dictation) was not used by any of the instructors. Peer-assessment was used by two of the instructors, while self-assessment was not used by any. Three out of five instructors used a standardized speaking test provided with the textbook materials.

Finally, in relation to the procedures of assessment and evaluation, three out of the five instructors used items from published textbooks, while one used items developed by himself, and another used items found on the Internet. As for providing feedback to students, all of the instructors told the total test score to the students and provided written comments, while two out of five provided verbal feedback and gave a checklist to the students showing them what they did or did not do and how well it was or was not done. The same was done with the final report.

The interview revealed some very interesting comments from the instructors involved. The first question ("When I say the word assessment, what images or feelings does the word conjure?") revealed that four out of five of the instructors' assessments have a negative meaning, especially among students and when they themselves have had to take assessments. Words such as "dread" (Peter) and "fear" (Anne) came up but also words like "necessary" (Peter, Robert) and "needed to sort out the good from the bad" (Peter, Rodney). All of the instructors felt that the assessment of their students was "going well" or "as far as I can tell everything is working out so far" (Rodney). In terms of what they felt most comfortable assessing, "pronunciation" (Peter, Anne) and "grammar" (Peter, Rodney, Carrie) were often mentioned. They were all also comfortable about the last assessment they did.

In the planning stage of the interview, the instructors who used the textbook materials for the assessment said things such as "it was convenient and relevant to use them" (Peter), and "they seemed appropriate for the level of the students" (Anne), while the instructors who used their own materials said that "I prefer to tailor my materials for the class to better meet their needs" (Rodney), and the one who used the Internet said that "I found a good site that provides good materials" (Carrie). All of the instructors said that they had complete freedom in choosing the kind of assessment for the class, although they were aware it should be relevant in some way to what they were teaching about. As for how they knew that the assessments would be suitable for their students, things such as "previous experience from doing this before" (Anne), "I've been teaching Japanese students for a long time" (Peter), and "the students seemed to find it reasonable and coped OK" (Rodney) came up.

Four out of the five instructors said they prepared the students for the assessment by telling them that it would be "based on the grammar and vocabulary of the unit studied" (Peter, Anne, Rodney, Carrie). In terms of how the students were able to approach the

assessment, four out of the five instructors said that the students were given questions to ask – some of which were required and some could be chosen from a list – while the other instructor allowed the students to talk about a topic they were interested in but told to use some of the grammar and vocabulary discussed in or related to what was studied in class for that particular unit. Three out of five instructors explained that they created their own standards and criteria for the oral assessment, while two of the five made their own standards to use in the assessment.

Some notable answers in the implementation stage of the interview came up in relation to whether the instructors felt that they were giving feedback to the students during the assessment. “I tried to do so with gestures and eye contact” (Peter) and “I didn’t want to give too much as it could affect the outcome” (Carrie) came up. All of the instructors felt that the students were able to complete the assessment to the best of their abilities, and none of them felt that they would change anything if they were to do it again as, as one pointed out, “I’ve been doing this for a long time already and haven’t had any major problems” (Rodney).

At the monitoring stage of the interview, all of the instructors rated the assessment as “fine” and “relevant”, and will use the results to help them to calculate the final grade of the students in the class. All of the instructors provided some form of written feedback after the test. Only the instructor and the student will have access to the results of the test in the case of all of the instructors. All of the instructors felt that the effects of the assessment were mostly positive as it allowed students to speak and the instructor had a clear opportunity to check their speaking levels. They were also united in that the result of the assessment would affect students differently depending on their performance in the interview as students appear to know if they have done particularly well or badly.

In the final dissemination stage (questions 18 to 23) of the interview, all of the instructors felt that the assessment was effective in judging the achievement of the students, although two said that “time restrictions” limited the effectiveness as they did not want to spend too much time on the assessment when instead the time could be spent doing actual teaching. No restrictions are in place at the university to affect how many students can achieve certain grades, and so the instructors have complete freedom to decide what grades the students eventually receive.

While all of the instructors provided feedback to students after their assessments, there was not much feedback taking place during the assessment. This may have been done for several reasons including time restraints and not wanting to unfairly affect the outcome of an assessment, but this is a chance wasted for positive washback as students who receive feedback immediately are more likely to remember it and use it to improve the remainder of their performance or one takes place in the future (McKinley, & Thompson, 2018). In my experience, in particular, students are often interested in little more than their final score, and this was also mentioned by three out of the five instructors during the interviews conducted.

Instead of merely providing students with written feedback, it might be more useful if verbal feedback was provided either during or after the assessments depending on the kind of oral assessment being done. If the assessment is a presentation, for example, then feedback after and not during would be more appropriate to avoid the student being embarrassed during the class, and peer-assessment should be encouraged based on the literature review which suggests it can be valuable and little bias occurs in most situations. Self-assessments may be useful as a way of seeing if a student understands how their

performance was, but it appears to be unhelpful when used as part of the actual assessment method.

## 5. Conclusion

This study was limited to five native-English instructors at one university in Japan due to the complicated nature of conducting and transcribing interviews. More studies involving other instructors in similar or different situations would help to add more knowledge and insight into the issues surrounding teacher-based assessment of oral assessments in Japan and other countries. Since the analysis was restricted to five participants, the results cannot be generalized. This research should be expanded to include a larger number of participants in order to generalize the results. Furthermore, this analysis lacks classroom data and does not provide the voices of students (one of the test's key stakeholders). In other words, additional classroom data gathered through observation and teacher interviews would have been beneficial.

Cultural and administrative issues may affect what an instructor can or cannot do in certain teaching environments. More research on how to ensure as much positive washback as possible in oral assessments would enable teachers to ensure that assessments are seen more positively and are also used as a way of improving English speaking levels. This analysis, on the other hand, may be used as a benchmark, providing guidance for future research. It is hoped that the results would add to the current test of washback literature.

In this study it has been shown that all of the instructors appeared to understand the importance of trying to promote positive washback, but more needs to be done to ensure that every aspect is covered. Providing more feedback – and at relevant times – depending on the kind of oral assessment being undertaken must be carefully considered and appropriate. Through more mixed method studies of this nature we can continue to learn more about how to provide positive washback with oral assessments in each particular setting and language school or university in each country. Further research in the field of language testing and assessment is suggested to draw on the results of the study and to delve deeper into the factors that influence the essence of washback and how instructors feel about evaluation and assessment.

## References

- Alderson, C. J., & Wall, D. (1993). *Does Washback Exist? Applied Linguistics*, 14 (1), 115–129.
- Azmi, M. N. L., Sham, R. M. (2018). *The Washback Effect of School-Based Oral Evaluation Performance and Its Causal Relationship with Fear of Negative Evaluation*. *Language Literacy*, Volume 2, Number 1, June 2018. Retrieved from [https://www.academia.edu/37078380/he\\_Washback\\_Effect\\_of\\_School\\_Based\\_Oral\\_Evaluation\\_Performance\\_and\\_Its\\_Causal\\_Relationship\\_with\\_Fear\\_of\\_Negative\\_Evaluation?email\\_work\\_card=title](https://www.academia.edu/37078380/he_Washback_Effect_of_School_Based_Oral_Evaluation_Performance_and_Its_Causal_Relationship_with_Fear_of_Negative_Evaluation?email_work_card=title)
- Cheng, L., Rogers, T., & Hu, H. (2004). *ESL/EFL Instructors' Classroom Assessment Practices: Purposes, Methods and Procedures*. *Language Testing*, 21(3), 360-389.
- Davison, C., & Leung, C. (2009). *Current Issues in English Language Teacher-Based Assessment*. *TESOL QUARTERLY* Vol. 43, No. 3, September 2009.
- Dawadi, S. (2019). *Students' and Parents' Attitudes Towards the SEE English Test*. *Journal of NELTA*, 1-2, 1-16.
- Estaji, M., Ghiasvand, F. (2019). *The Washback Effect of IELTS Examination on ELF Teachers' Perceived Sense of Professional Identity: Does IELTS Related Experience Make a*



- Difference?* Journal of Modern Research in English Language Studies 6(3), 83-103. Retrieved from [https://www.academia.edu/40839168/The\\_Washback\\_Effect\\_of\\_IELTS\\_Examination\\_on\\_EFL\\_Teachers\\_Perceived\\_Sense\\_of\\_Professional\\_Identity\\_Does\\_IELTS\\_Related\\_Experience\\_Make\\_a\\_Difference?email\\_work\\_card=title](https://www.academia.edu/40839168/The_Washback_Effect_of_IELTS_Examination_on_EFL_Teachers_Perceived_Sense_of_Professional_Identity_Does_IELTS_Related_Experience_Make_a_Difference?email_work_card=title)
- Green, A. (2013). *Washback in Language Assessment*. International Journal of English Studies, 13 (2), 39–51.
- Hazaca, A. N., Tayeb Y. A. (2018). *Washback Effect of LOBELA on EFL Teaching at Preparatory Year of Najran University*. International Journal of Educational Investigations. Vol. 5, No.3: 1-14, June 2018. Retrieved from [https://www.academia.edu/37625472/Washback\\_Effect\\_of\\_LOBELA\\_on\\_EFL\\_Teaching\\_at\\_Preparatory\\_Year\\_of\\_Najran\\_University?email\\_work\\_card=view-paper](https://www.academia.edu/37625472/Washback_Effect_of_LOBELA_on_EFL_Teaching_at_Preparatory_Year_of_Najran_University?email_work_card=view-paper)
- Hughes, A. (2003). *Testing for Language Teachers*. Cambridge, New York: Cambridge University Press.
- Khan, A. B. M. A., Aziz M. S. A. A., & Stapa S. H. (2019). *Examining the Factors Mediating the Intended Washback of the English Language School-Based Assessment: Pre-service ESL Teachers' Accounts*. Retrieved from [http://journals-jd.upm.edu.my/Pertanika%20PAPERS/JSSH%20Vol.%2027%20\(1\)%20Mar.%202019/4%20JSSH-2622-2017.pdf](http://journals-jd.upm.edu.my/Pertanika%20PAPERS/JSSH%20Vol.%2027%20(1)%20Mar.%202019/4%20JSSH-2622-2017.pdf)
- Kim, Y. (2009). *An investigation into Native and Non-Native Teachers' Judgments of Oral English Performance: A mixed Methods Approach*. Language Testing 26(2):187-217
- Loumbourdi, L. (2014). *The Power and Impact of Standardised Test: Investigating the Washback of Language Exams in Greece*. Frankfurt am Main: Peter Lang.
- Matsuno, S. (2009). *Self-, Peer-, and Teacher-Assessments in Japanese university EFL writing classrooms*. Language Testing, 26(1), 075-100 – January 2009
- McKinley, J., & Thompson, G. (2018). *Washback Effect in Teaching English as an International Language*. *TESOL Encyclopedia of English Language Teaching*. 1. 2018.
- Muñoz, A. P., & Álvarez, M.E. (2010). *Washback of an Oral Assessment System in the EFL Classroom*. Language Testing 2010; 27; 33
- Prodromou, L. (1995). *The Backwash Effect: From Testing to Teaching*. *ELT Journal*, 49(1), 13–25.
- Saif, S. (2006). *Aiming for Positive Washback: a Case Study of International Teaching Assistants*. Language Testing, 23 (1), 1–34.
- Saville, N. D. (2009). *Developing a Model for Investigating the Impact of Language Assessment within Educational Contexts by a Public Examination Provider*. An unpublished PhD thesis, University of Bedfordshire, England.
- Sugimoto, Y. (2014). *An Introduction to Japanese Society* (4<sup>th</sup> ed.). Cambridge: Cambridge University Press.
- Takagi, A. (2010). *A Critical Analysis of English Language Entrance Examinations at Japanese Universities*. An unpublished PhD thesis, University of Exeter, England.
- Tsagari, D. (2012). *FCE-exam Preparation Discourses: Insights Form an Ethnographic Study*. *ESOL Research notes* 47.
- Wicking, P. (2017). *The Assessment Beliefs and Practices of English Teachers in Japanese Universities*. *JLTA Journal*, vol. 20: pp. 76-89, 2017.
- Xie, Q. (2013). Does Test preparation work? Implications for score validity. *Language Assessment Quarterly*, 10 (2), 196–218.