

Voices in the Field: An Interview with Yoshinori Watanabe

by Tim Newfields

Yoshinori Watanabe is an associate professor in the Faculty of Education and Human Studies at Akita University. He received his Ph.D. from Lancaster University in the UK. He has also been involved in a number of research projects, including the investigation of curricular innovations commissioned by the MEXT Ministry, a comparative study of English abilities of Japanese, Korean and Chinese high school students, and so forth. In 2004 he coedited Washback in Language Testing with Liying Cheng and Andrew Curtis. This interview was conducted electronically in May 2005.

How did you become interested in language testing?

At the beginning of my ESL/EFL research career, I was interested in linguistic analyses of learner errors. This topic had subsequently developed into a series of studies on learner strategies. Meanwhile, I became curious to know how students' language strategies varied in test-taking and non-test-taking situations. This led to an interest in how language tests could best be constructed to help students learn and teachers teach. Now I place a greater focus on tests than on strategies, and yet strategies are still on my research agenda as one of the factors that are influenced by tests.

How has your understanding of washback changed over the years?

When first embarking on washback research, I assumed that innovation in tests would automatically lead to innovation in teaching and learning. But now I have to admit that this view was overly naive. It is certainly important to construct a good test, but that is not sufficient to improve teaching or learning. Even if a test is good, if we don't know how to make the best use of it, it will not result in beneficial washback. So I think teachers should not only know how to make good tests, but also how to appropriately apply test content in class. In a sense, teachers can be viewed as agents for educational change via the tests they use. Putting blame on tests won't get us anywhere: we should "make friends" with tests.

The concept of "critical language testing" is briefly mentioned in the recent text you co-edited. Can you explain what is meant by that concept?

The core idea of critical language testing can be characterized by the constant monitoring of test use. No test is used in a vacuum. Any test is used to serve for specific purposes. Language tests are no exception. So it's important part of research work to examine empirically how tests are used in a given context, and report to the general public whether the test has successfully achieved the goal. A traditional view of language testing tended to view language tests in a narrow technical sense, restricting its focus to establishing reliability and validity. But if we want to examine educational uses of language tests, this kind of narrow definition won't work. All we can do by this approach is just to infer effects of tests. However, in critical language testing view, it is argued that we have to take into account not only educational contexts where the test used, but even wider societal contexts where the results of tests are put to use.

You mentioned that the mass media often tends to simplify and exaggerate how tests impact actual classroom practices. Why do you feel this is so?

When I mentioned problems with the mass media, I indirectly criticized policy makers. My research suggests that they tend to promote educational changes based on general opinions expressed in the mass media rather than on empirical evidence. The majority of opinions regarding entrance exams, for example, focus on its negative effects. So every few decades, there's an argument that the English component should be eliminated from the exam. But the rationale for that is very weak, because it's not based on empirical evidence. I don't think making educational policies based on public opinions expressed in the mass media is productive since opinions are easily manipulated from day to day, and rarely depict what is happening accurately.

What do you feel have been the main limitations with most washback research studies conducted so far?

Many good research studies have recently begun to appear in international journals. However, I think there're still a number of research studies that are not fully empirical. The results of such studies are based on many unproven assumptions. For example, it does not help much to examine test papers in order to analyze washback. Analysis of test papers might illuminate other phenomena, but not washback. For example, it has been established empirically that paper and pencil tests of pronunciation do not provide a valid measurement of a student's pronunciation ability (Buck, 1989). Even if a person scores high on such a test, it does not guarantee that he/she will pronounce well. Merely looking at such a test provides no idea of how students are preparing for it, or how teachers supposedly teach it. In order to examine washback effects, we must observe teaching and learning behaviors. I think good research should be empirical in nature. Just exchanging opinions does not much help us understand much about washback. Empirical research should provide a number of key details which readers may confirm: where and how the data were gathered, what type of data was gathered and so on. I think also it's important to build a model based on empirical research studies results at some point, to make the information more meaningful and provide guidelines for future studies.

In order to generate better positive test washback, you recommend "re-attribution training" for teachers and students. Could you describe this concept a bit more?

The research results indicate so far that washback is a highly psychological phenomenon and not something inherent in any test. One of the biggest factors that hinders positive washback from being generated is the negative attitudes teachers and learners hold towards tests. Many of them feel tests are a necessary evil. Worse than that, teachers and learners are prone to attribute what they can't do to tests. Many love to claim, "because there's a test, we have to study grammar," or "because there's a test, we can't teach communication," and so on. As long as such negative attitudes persist, it's not possible to generate positive washback, no matter what test is used. Interestingly, such complaints seldom accurately reflect the tasks and contents of the target tests. Unless we overcome this way of thinking, positive washback can never be achieved. Re-attribution training is one method to help test users change their beliefs about tests. I won't go into details of theories here, but I hope one example will suffice. Suppose your student scored very low in a test. If he believes that this was due to a poor language learning ability, that student is unlikely to study hard for the next test. One thing that you could do as a teacher is to help him to attribute the cause of a poor score to something he has control over. For example, you may encourage him to realize the problem is not his learning ability, but rather his basic study method. More about this is available in *Learned helplessness with children: Prevention and alleviation through success experiences or re-attribution training* (1978) by H. E. Harris.

What sort of future research on washback would you like to see conducted?

To date most of the research has been concerned with so-called high-stakes tests. I would like to see more washback studies of classroom-based tests conducted. If we can discover how to motivate learners through tests this would be very helpful. A number of questions should also be explored in this regard. Do short pop quizzes motivate learners? When and how often should in-class quizzes be carried out? What sort of feedback is most useful for students? How should announcements be made to students prior to a test? Good teachers should perhaps have an intuitive sense of these answers. However, research is needed to confirm or disprove intuitions and suggest ways of exploiting positive test washback.

What sort of projects are you working on currently?

One of the projects I am currently involved in concerns how to motivate learners and teachers by using tests. As the book *Washback in Language Testing* attests, the presence of test does not necessarily drive learners and teachers to work harder than otherwise. I've been examining the conditions under which a test motivates learners as well as features of tests that are likely to induce learners to learn harder.

Another topic I've just started is the validity of teacher exams. I've long doubted that teacher examinations that are being conducted in Japan really measure the knowledge and skills that teachers need to use at school settings. It's been recommended by the MEXT Ministry that English teachers should achieve a certain level of scores on standardized tests, such as the TOEFL, TOEIC, and STEP. However, these tests are not intended to measure specific types of ability that are required of English teachers. I just started to carry out needs analysis, and will eventually develop a model of teacher exams.

Works Cited

Buck, G. (1989). Written tests of pronunciation: do they work? *ELT Journal*, 43 (1), 50-56.

Harris, H. E. (1978). *Learned helplessness with children: Prevention and alleviation through success experiences or re-attribution training*. Winston-Salem, NC: Wake Forest University.

Cheng, L., Watanabe, Y., Curtis, A. (Eds.) (2004). *Washback in language testing: Research contexts and methods*. Mahwah, N.J.: Laurence Erlbaum & Associates.

HTML: http://jalt.org/test/wat_new.htm /

PDF: <http://jalt.org/test/PDF/Watanabe.pdf>