Test Washback and Implications for TOEIC© Course Syllabus Design

Nicholas A. Caine

This two-part paper presents the notions of *test washback* and *test impact* and looks at the implications that the two may have for the design of TOEIC courses at university level in Japan. In Part 1, a detailed overview of the terms *washback* and *impact* is offered before the discussion moves on in Part 2 to consider the possible effects that the TOEIC may have upon teaching and learning. Whilst much of the research in this area suggests an overall negative effect (e.g. Buck 2001; Thomson 2012; Newfields 2005; Irwin & Nagy 2010), the extent to which *positive* washback can be nurtured from the TOEIC will be addressed in Part 2 when some practical suggestions for the language classroom will also be presented.

Background

Developed by the US-based Educational Testing Service (ETS), the Test of English for International Communications™ (TOEIC©) was first implemented in Japan in 1979, taking the form of a paper-based multiple-choice listening and reading test. Although a limitation of the TOEIC in its original form was that it could only measure the productive skills of speak-

ing and writing *indirectly*, the format remained relatively unchanged until a major revision was announced in July 2005. As part of these changes, an additional TOEIC Speaking and Writing Test was added to the ETS suite from 2006. Despite the introduction of this additional test, the TOEIC Listening and Reading Test remains at the forefront of those offered by ETS and may even help to subsidize the less popular TOEIC brand tests (McCrostie 2010).

In Japan, where the TOEIC is used not only by companies for hiring and training decisions, but also by universities who have included it as a core part of the English curriculum (in addition to using it for placement purposes), it is important to be aware of the washback and wider societal effects of the test. Many educators may feel a certain degree of frustration at being unable to exert influence over the content or format of the test, but in view of the ubiquitousness of the TOEIC in Japan, we should try to tap into its potential for positive *washback*. Before looking at washback in specific relation to university TOEIC courses, however, the proceeding discussion here in part one reviews some of the key ideas and findings within the area of test *washback* and *impact* in order to provide the necessary background to part two of the discussion.

What is test washback？

Various explanations of the term ‘ washback ’ can be found throughout the published research and literature on language testing. One of the most common definitions sees the concept referred to as the *influence* of testing on teaching and learning (Alderson & Wall 1993; Gates 1995; Cheng & Curtis 2004). Similarly, Shohamy et al (1996: 298) define washback as ‘ the connections between testing and learning ’ and Saville (2000: 4)

and Hughes[1] (1989: 1) as ' the effect of testing on teaching and learning '. Messick (1996: 241), noting that washback can have either harmful or positive effects, describes it as ' the extent to which the introduction and use of a test influences language teachers and learners to do things they would not otherwise do that promote or inhibit language learning '.

Such definitions refer only to washback in terms of the influence that tests might have upon the language classroom and the participant roles of teacher and learner. In other analyses a much broader interpretation is offered, taking the view that tests can have more far-reaching effects within the field of education. For example, Andrews (2004: 37), in an article that explores the relationship between washback and curricular innovation, looks beyond the classroom, and uses the term to describe ' the effects of tests on teaching and learning, the educational system, and the various stakeholders in the education process '. The view of washback presented here approaches what some writers differentiate as *test impact*.

What is test impact?

Although the terms *washback* and *impact* are sometimes used synonymously—as indeed demonstrated by Andrew's (2004) definition of the former, above—test impact more accurately refers to the wider implications and effects of a given test. For example, whilst acknowledging that washback can be seen as the *influence* and *effect* of tests on teaching and learning, McNamara (2000: 74) notes the wider effect that tests can exert

1 Whilst ' washback ' is the preferred term in British applied linguistics (Alderson & Wall, 1993: 115), some writers such as Hughes refer to the same phenomenon as ' backwash '.

on the community as a whole.

Wall (1997: 291) makes a similar distinction between test *washback* and test *impact*, agreeing that the term ‘ impact ’ more accurately refers to ‘ any of the effects that a test may have on individuals, policies or practices, within the classroom, the school, the educational system or society as a whole ’.

Referring to the wider societal effects of a test (i.e. test impact), Rea-Dickens (1997) expands a little further by identifying ‘ at least 5 stakeholder categories: learners, teachers, parents, government and official bodies, and the marketplace ’. This is certainly the case in Japan, particularly with regard to the latter, where prospective employers attach an increasing weight and importance to candidates ’ TOEIC scores.

Bailey (1996: 263-264) also adopts a more holistic view regarding the effects of tests, but prefers to consider overall *impact* in terms of ‘ washback to the learners ’ and ‘ washback to the programme ’—the latter group including ‘ teachers, administrators, curriculum developers, counsellors, etc. ’ (Bailey 1996: 264).

It becomes clear then that the term ‘ washback ’ is open to a variety of interpretations and that there are a number of important variables to consider when conducting research into the issue. For example, how are we defining the term ‘ washback ’ exactly? Are we using the term to describe the effects of a test on teaching and learning only, or are we using it in the wider sense to include the effects of a test on other stakeholders in the education process, which as noted above would be more accurately referred to as test impact?

At this point, let us look at the Washback Hypothesis (Alderson & Wall 1993) in order to consider how a test might affect—or *influence*—the language teaching and learning environment.

The Washback Hypothesis

The title of an article written by Alderson & Wall (1993) posed the following question: " Does Washback Exist ? " In the article, it was noted that whilst many assertions had been made relating to the influence of tests, there was very little in the way of empirical evidence to back up the claims that tests affected teaching and learning, in either a negative or positive way. In order to examine in greater depth the relationship between washback and ' influence ', Alderson & Wall (1993: 120-1) put forward the 15 hypotheses listed below, highlighting more specifically some of the ways in which a test might affect teaching and learning. Bailey (1996: 265-266) notes that five of the hypotheses relate specifically to ' washback to the learners ' and six relate to ' washback to the programme '. Thus, the letters in parentheses have been added in order to illustrate this point, with ' L ' indicating those statements concerning washback to the learners, ' P ' to those concerning washback to the programme.

Some possible Washback Hypotheses

(1) *A test will influence teaching.* (**P**)
(2) *A test will influence learning.* (**L**)
(3) *A test will influence* ***what*** *teachers teach* (**P**); *and*
(4) *A test will influence* ***how*** *teachers teach* (**P**); *and therefore by extension from* (2) *above*:
(5) *A test will influence* ***what*** *learners learn* (**L**); *and*
(6) *A test will influence* ***how*** *learners learn.* (**L**)
(7) *A test will influence the* ***rate*** *and* ***sequence*** *of teaching* (**P**); *and*
(8) *A test will influence the* ***rate*** *and* ***sequence*** *of learning.* (**L**)

(9) *A test will influence the* ***degree*** *and* ***depth*** *of teaching* (**P**); *and*

(10) *A test will influence the* ***degree*** *and* ***depth*** *of learning.* (**L**)

(11) *A test will influence attitudes to the content, method, etc. of teaching and learning* (**P**)

(12) *Tests that have important consequences will have washback*; *and conversely*

(13) *Tests that do not have important consequences will have no washback*.

(14) *Tests will have washback on* ***all*** *learners and teachers*.

(15) *Tests will have washback effects for* ***some*** *learners and* ***some*** *teachers, but* ***not*** *for others*.

We should also take into account that even a ‘ poor ’ test could have a positive washback effect if it encouraged motivation on the part of learners or teachers. For example, a test might encourage learners to ‘ do their homework, take the subject being tested more seriously, and so on ’, whereas teachers might ‘ prepare lessons more thoroughly ’ (Alderson & Wall 1993: 117). However, if teachers are unduly influenced by a test this could have serious implications if the content of that test fails to reflect the aims and objectives of a recommended course of study or prescribed syllabus—the result being that significant areas of potentially enriching course content which are not tested may then be seen as irrelevant. Consequently, this content is neglected and, in some cases, results in a ‘ narrowing of the curriculum ’ (Madaus 1988, *cited in* Wall 2000: 500).

Washback as a result of ‘ high stakes ’ testing

According to the‘ Washback Hypothesis ’(Alderson & Wall 1993: 120-1)

important tests are likely to have washback. In other words, tests will arguably have a greater influence on teaching and learning in a' high-stakes ' situation—i.e. one in which the test is typically used ' to compare and rank individuals, schools or national systems ' (Chapman & Snyder Jr. 2000: 458) and whose ' primary use is to ration future opportunity as the basis for determining admission to the next layer of education or to employment opportunities '(*ibid*.).

The research findings of studies that have involved so-called ' high-stakes ' tests confirm some evidence of washback, although it is often reported that, in the case of teaching, such tests have a greater effect upon *content* and rather less upon the actual methodology employed by teachers. For Example, Cheng (1997), who looked at the effect that changes to the HKCEE[2] in English had upon teaching and learning, found that washback occurred, but only in terms of bringing about change to teaching materials. Evidence that the changes to the test brought about changes to the *way* teachers taught was not conclusive.

Likewise, in the context of Japan, research conducted by Watanabe (1996; 2004) examining the belief that teachers' reliance on grammar-translation comes about as a direct result of university examination content, shows that the examinations exert a washback effect on some teachers but not on others. His findings suggest that factors such as the educational background, personal beliefs and experience of the individual teacher have as much influence over which methodology is employed in the classroom (Watanabe 1996: 330-1).

Chapman & Snyder Jr. (2000: 462) also question the extent to which high-stakes testing influences teachers' classroom methodology. Citing a

2 The Hong Kong Certificate of Education Examination

general education study from Uganda by Snyder et al. (1997), which found that changes made to a national examination did not have the desired effect of encouraging teachers to alter their instructional practices, they suggest that ‘ it is not the examination itself that influences teachers' behavior, but teachers' beliefs about those changes ’ (Chapman & Snyder Jr. 2000: 462).

Conclusion

It is generally accepted, then, that all tests are likely to influence teaching and learning to some degree, whether it be in the form of positive washback or negative washback, but evidence that illustrates the extent to which tests influence content, methodology (or both) is not always conclusive. With a “ high-stakes ” test such as the TOEIC we are likely to see the occurrence of both *washback* and *impact*. Indeed, the fact that this test is being used for purposes other than those for which it was originally designed perhaps causes the most controversy amongst educators in Japan. However, whilst the merits and demerits of the TOEIC provoke much discussion within the field of English education here it would be churlish to dismiss it out of hand as having a predominantly negative effect on classroom teaching and learning. While the scope of this paper is unable to cover in sufficient depth the *extent* of TOEIC washback or impact in relation to teaching methodologies or learning preferences in the chosen context—that should be reserved for a future longitudinal study that is able to include classroom observation in addition to questionnaires, interviews, surveys etc—it is hoped that Part 2 will illustrate the ways in which the TOEIC influences (or *should* influence) syllabus *content*. It will attempt to shed light on some of the positive aspects of the test and argues

that lesson content should go beyond simply practising and reviewing test items.

References

Andrews S (2004) ' Washback and Curriculum Innovation ' *in* Cheng L, Y Watanabe & A Curtis (Eds.)

Alderson J C & D Wall (1993)' Does washback exist? '***Applied Linguistics*** 14/2: 115-129

Bailey K M (1996) ' Working for washback: a review of the washback concept in language testing ' ***Language Testing*** 13/3: 257-279

Brown J D & S O Yamashita (Eds.) (1995) ***Language Testing in Japan*** Tokyo: Japan Association for Language Teaching

Buck G (2001) ***Assessing Listening*** Cambridge: CUP

Chapman D W & C W Snyder Jr (2000) ' Can high stakes national testing improve instruction: reexamining conventional wisdom ' ***International Journal of Educational Development*** 20: 457-474

Cheng L (1997) How does washback influence teaching? Implications for Hong Kong ' ***Language and Education*** 11/1: 38-54

Cheng L & A Curtis (2004) ' Washback or backwash: A review of the impact of testing on teaching and learning ' *in* Cheng L, Y Watanabe & A Curtis (Eds.)

Cheng L, Y Watanabe *with* A Curtis (Eds.) (2004) ***Washback in Language Testing: Research Contexts and Methods*** New Jersey: LEA

Clapham C & D Corson (Eds.) (1997) ***Encyclopedia of Language and Education: Vol.7. Language Testing and Assessment*** Dordrecht: Kluwer Academic

Gates S (1995) ' Exploiting washback from standardized tests ' *in* Brown J D & S O Yamashita (Eds.)

Hughes A (1989) ***Testing for Language Teachers*** Cambridge: CUP

Irwin & Nagy (2011) ***Shiken: JALT Testing and Evaluation SIG Newsletter*** 15/2: 10-19 Retrieved November 27, 2013 from http://jalt.org/test/PDF/Irwin&Nagy.pdf

McCrostie J (2010) ***Shiken: JALT Testing and Evaluation SIG Newsletter*** 14/1: 2-10 Retrieved November 27, 2013 from http://jalt.org/test/PDF/McCrostie1.pdf

McNamara T (2000) ***Language Testing*** Oxford: OUP

Messick S (1996) ‘ Validity and washback in language testing ’ ***Language Testing*** 13/3: 241-256

Newfields T (2005) ‘ TOEIC washback effects on teachers: A pilot study at one university faculty ’ ***Toyo University Keizai Ronshu*** 31/1: 83-106 Retrieved November 27, 2013 from http://www.tnewfields.info/Articles/PDF/Newfields-Washback.pdf

Rea-Dickins P (1997) ‘ So, why do we need relationships with stakeholders in language testing? A view from the UK ’ ***Language Testing*** 14/3: 304-314

Saville N (2000) ‘ Investigating the impact of international language examinations ’ ***University of Cambridge Local Examinations Syndicate*** (***UCLES***) ***Research Notes*** *2* Retrieved November 27, 2013 from http://www.cambridgeenglish.org/images/22642-research-notes-02.pdf

Shohamy E, S Donitsa-Schmidt & I Ferman (1996) ‘ Test impact revisited: Washback effect over time ’ ***Language Testing*** 13/3: 298-317

Snyder C W Jr, B Prince, G Johanson, C Odaet, L Jaji & M Beatty (1997) ‘ Exam Fervor and Fever: Case Studies of the Influence of Primary Leaving Examinations on Uganda Classrooms, Teachers and Pupils (Volume One) ’ ***Advancing Basic Education and Literacy*** *(****ABEL****)* ***Project*** Academy for Educational Development, Washington DC

Thomson S (2012)‘ The Effects of TOEIC Education in South Korean Universities ’ Retrieved November 27, 2013 from http://www.birmingham.ac.uk/Documents/college-artslaw/cels/essays/matefltesl dissertations/Thomson-tefl--diss.pdf

Wall D (1997) ‘ Impact and washback in language testing ’ *in* Clapham C & D Corson(Eds.)

Wall D(2000) ‘ The impact of high-stakes testing on teaching and learning: can this be predicted or controlled? ’ ***System*** 28/4: 499-509

Watanabe Y (1996) ‘ Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research ’ ***Language Testing*** 13/3: 318-333