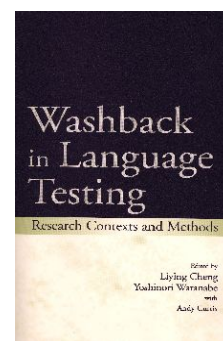


Book Review

Washback in Language Testing: Research Contexts and Methods

Edited by Liying Cheng and Yoshinori Watanabe, with Andy Curtis (2004)

Mahwah, N.J.: Laurence Erlbaum & Associates (237 + xxi pages)



How do language tests influence students, teachers, and other test stakeholders? What are the best ways to measure the extent of that influence? This volume addresses both questions. Its eleven chapters summarize many significant language washback studies to date and suggest directions for further research.

The book opens with a historical overview of washback and various attempts to define it. Though washback is often described as "the influence of testing on teaching and learning" (Anderson & Wall, 1993), other researchers have underscored different dimensions of this concept. Though washback is neither intrinsically positive or negative, most tests have both positive and negative results.

Advice for those undertaking washback research from ethnographic or qualitative perspectives is the focus of the second chapter. The value of using multiple measures to ascertain test impact was underscored, along with the need for researchers to clearly state their own baseline biases.

The book then explores the relationship between curricular reform and testing. The author cautions against "the naive reliance of high-stakes tests as a primary change strategy" (p. 48) and recommends that test designers take a sober look at the large number of educational innovations that hit the rocks because the limits of coercive testing to induce long-term change were not adequately acknowledged. The importance of being aware of antecedent conditions and "process factors" (Rogers, 1983) which facilitate or impede any educational reforms is duly underscored. Noting how even carefully planned tests can have unexpected consequences or incur resistance, the author reminds test designers to involve test stakeholders in all test design phases. Ways that publicized claims about washback often differ from research evidence is pointed out. In particular, the claim that any given test will invariably produce positive curricular change needs to be viewed with caution. Given the climate of increasingly measurement-driven instruction in many parts of the world, the caveats against over-testing and undue emphasis on surface learning seem timely.

The effects of standards-based, assessment-driven testing on writing instruction in one American state is investigated in Chapter 4. Self-reports by teachers and school principals suggest curricular focus had significantly changed to match state-mandated test priorities. Noting that, "Assessment plays a key role in signaling priorities among standards and making student performance expectations concrete," (p. 68) the authors concede that essay writing skills may have improved as a result of the new exam, but standards which were not being tested have become increasingly ignored. The point exemplifies the seemingly never-ending conundrum over the negative impact of testing on educational objectives.

The next chapter describes the development of a rating form to ascertain how the IELTS might be impacting EFL texts. The most interesting thing in this chapter was the comment about the distinction between piloting and true validation. Though many research studies are piloted, few (included the one

highlighted in this chapter) are actually fully validated: validation is a rigorous and time-consuming process.

The impact of the IELTS on two academic courses is examined next. A course which focused almost exclusively on preparation for the IELTS test was compared with another which focused on more general academic skills. The former class tended to be teacher-centered and use a more restricted range of teaching materials than the wider ESP class. However, except for a slight improvement in the listening section of the IELTS in the mainly test-centered class, there was no statistically significant difference in IELTS scores among students from either class after a period of 3-4 weeks. To gauge how test-oriented and non-test-oriented classes differ, longer longitudinal time frames are probably needed. One interesting insight from this chapter was information concerning how IELTS-related skills could be taught a number of ways: over-reliance on mock tests is by no means the only way to prepare for a test.

Highlighting how teacher beliefs powerfully impact reactions to top-down educational change, the next chapter discusses Australian EFL teachers' reactions to a nationally mandated test. Comparing a pre-mandated test era sample of teachers with a post-mandated era sample, self-reports did not suggest any statistically significant differences in the way teachers taught during that 8-year time span, though the content of many classes likely did change. The data underscored how washback effects vary from teacher to teacher: some teachers adopt new testing practices quite readily, while others are less willing to change.

The next chapter explores the effects of university entrance exams on secondary school instruction in Japan. Observing five teachers from three high schools, the author corroborates a premise echoed throughout this volume: that test content is simply one of many variables influencing educational change. Stressing that "innovation in testing does not automatically bring about improvement in education." (p. 130), Watanabe suggests that positive washback is more likely to occur if teachers are familiar with a wide range of teaching methods, the test is perceived as having a high degree of face validity, and some form of re-attribution training (William & Burden, 1997) is provided. To promote positive backwash, Watanabe also recommends in-service teacher training and action research. However, systematic studies about each of these needs to be conducted before we are able to state with confidence what factors promote positive washback.

Liyong Cheng examines the influence of a new examination in Hong Kong on teachers in Chapter 9. Her results suggest examinations tend to have a more direct impact on course content than methodology. Though most teacher behaviors did not change significantly as a result of a new test introduced in Hong Kong, the attention to classroom content and focus on homework did. Cheng's comment about how "context and washback co-construct each other" (p. 148) was intriguing. One thing her article could explore in more detail is the macroscopic forces impelling curricular/test alignment. Other parts of this volume suggest there are political forces in motion which are compelling schools to use increasingly measurement-driven instruction. A question to be addressed more fully is why this might be so.

The extent that a high school entrance test in mainland China has been shaping actual classroom practices is considered in Chapter 10. What the author found was a partial mismatch between what the test constructors envisioned should happen in class and what many teachers feel compelled to teach. Due to the high-stakes nature of this test, linguistic knowledge rather than pragmatic use became the

focus of most classes. One particularly interesting observation was the way that the chronological and conceptual courses structures differ: what teachers conceptually think should be taught in a curriculum and what they actually end up teaching frequently vary.

The final chapter discusses the effects of a national oral matriculation test for Israeli high school students. The introduction of an oral component to the matriculation exam in 1986 has had mixed results. Positive effects include greater focus on oral skills in class and possibly higher oracy levels. Negative effects included curricular narrowing, heightened anxiety, as well as various forms of test score pollution. Both positive and negative washback were most pronounced among average students: markedly high achieving students were already confident of their ability to do well in the test, and low achieving students often felt pessimistic about the whole affair. The chapter concludes by discussing some unresolved ethical issues relating to testing: test designers have an ethical duty to consider the impact of exams on the educational system at large. Indeed, the whole notion of consequential validity (Messick, 1989) is designed to encourage this. Though many test designers spend a lot of time and money to increase the reliability and face validity of their exams, at least as much attention should be devoted to enhancing consequential validity.

Conclusion

The notion of washback has been discussed for at least three decades in the field of language testing. This volume successfully draws a diverse range of articles about test impact under one cover. Though many of the articles should be regarded as works in progress rather than final studies, this volume does effectively highlight some of the complexities involved in current washback research. It is my hope that a revised volume will be published in about a decade. By then we should be seeing new generation of washback studies which are increasingly sophisticated and refined.

- Reviewed by Tim Newfields
Toyo University

Works Cited

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 13-103). New York: Macmillan.

Rogers, E. M. (1983). *The diffusion of innovations*. (3rd edit). New York: MacMillian.

Williams, M. & Burden, R. (1997). *Psychology for language teachers: A social constructivist approach*. Cambridge, England: Cambridge University Press.