



Ministry of Higher Education and
Scientific Research
University of Tikrit
College of Education for Humanities
English Department
Ph.d Studies/ Methods of Teaching

Washback

Pro. Dr. Nagham Q. Yahya

nagyahy@tu.edu.iq

1.1 Introduction

The test is ‘high stakes’, defined by Madaus as a test ‘whose results are seen – rightly or wrongly – by students, teachers, administrators, parents, or the general public, as being used to make important decisions that immediately and directly affect them’ (Valette, 1988: 87), a phenomenon known as ‘washback’ may occur. Washback is the term that is used when students and teachers ‘do things they would not necessarily otherwise do because of the test’ (Alderson and Wall, 1993: 117). They might, for example, pay more attention to certain parts of the teaching syllabus at the expense of other parts because they believe these will be emphasised on the test.

There has been growing interest in washback over the last two decades, for theoretical, practical and political reasons.

- From theoretical reasons, Messick (1996) claimed that washback is a particular instance of the consequential aspect of construct validity, which suggests a corollary that investigating washback and other consequences is a crucial step in the process of test validation.
- The practical and political interest stems from the use that policy-makers make of high-stakes tests to influence educational practices.

The idea of washback takes on more complexity when we consider not only whether the effect of tests are positive or negative but also whether they are immediate or delayed, direct or indirect, or apparent or not visible – e.g. changes in attitude that do not manifest themselves in overt behaviour (Henrichsen, 1989: 80).

There are many accounts of the use of high-stakes testing in education and other realms of public life. Eckstein and Noah (1993: 5–17) discuss a number of functions that such tests have served in society: ending the monopoly over government jobs held by the privileged classes (e.g. competitive examinations in China during the Han Dynasty), checking patronage and corruption (the Indian Civil Service examination in nineteenth-century Britain), encouraging ‘higher levels of competence and knowledge’ (entry examinations to the professions in France and Germany), allocating sparse places in higher education (university entrance examinations in Japan), and measuring and improving the effectiveness of teachers and schools (the ‘Payment by Results’ system established in Britain in the 1860s, where state funding of schools was partly determined by the results students received in tests administered by

school inspectors). It is not difficult to imagine the influence that these tests might have had on the learning goals and methods of the candidates preparing for them. There is little empirical evidence available, however, to provide a link between these tests and the teaching and learning that are said to have resulted from them.

These changes were not to be seen as automatic, however. Improvements in educational standards depended not only on well-constructed tests representing the full range of educational objectives, but also on a clear articulation and exemplification of the standard desired, sensitive training of markers, and ample opportunity for teachers and students to understand and practice using the criteria for evaluating performance. Airasian (1988) was cautious in his views of whether tests could bring about changes in teaching and learning, emphasising the need to consider the level of cognitive skills assessed in the test and the likelihood that teachers could successfully train students in high-level operations such as reasoning, problem-solving and critical thinking. There are two different types of preparation activity on a scale from ethical to highly unethical . An example of an ethical practice was motivating students to study by discussing how important the test was to them. An example of unethical practice was using material in the classroom that was very similar to the material used in the test (see Mehrens and Kaminsky, 1989).

Alderson and Wall (1993) provided the first critical analysis of the notion of washback .They explained that washback was not the same as the general pressure caused by important tests – for example, the feeling that one should spend more time studying. The washback of a test was specific to that test alone, manifesting itself in decisions about how much attention to pay to certain aspects of the domain in question (e.g. the teaching syllabus), depending on the importance given to these aspects in the test.

They proposed a number of ‘washback hypotheses’ to illustrate the types of influences that a test could conceivably have:

- on what teachers taught and learners learned;
- on how teaching and learning took place;
- on the rate, sequence, degree and depth of teaching and learning;
- and on the attitudes of teachers and learners

Other early contributions to the understanding of washback came from Hughes (1994) and Bailey (1996, 1999).

Hughes introduced an important distinction between washback on the participants in an educational system, on processes and on products.

Bailey created a 'basic model' to illustrate the mechanism by which washback developed. She also proposed a series of questions to ask of any external. These questions probed the participants' understanding of the purpose of the test and how the results were to be used, the theoretical bases of the test, the use of authentic texts and tasks, the manner in which test results were provided, and other features she believed could influence the appearance and nature of washback.

Two discussion of the nature of washback :

- The first discussion was by Messick (1996), who first reviewed the notion of a unified version of construct validity, and then argued that washback was one (but not the only) manifestation of the consequential aspect of validity.
- The second discussion of the nature of washback was by Bailey (1996), who provided a comprehensive review of the literature on washback up to that point.

Four case studies provided the type of empirical evidence for washback that was often missing in publications before the 1990s:

- Alderson and Hamp-Lyons (1996) used teacher and student interviews and classroom observations to investigate the washback. They followed two teachers as they conducted 'normal' language lessons and TOEFL preparation lessons, and found that although there were differences between the two types of lessons for each teacher, there were differences that were at least as great between the two teachers themselves.
- Watanabe (1996) also designed a comparative study. He carried out observations and interviews with two teachers who were teaching preparation classes for one examination which emphasised grammar-translation and for one which did not. He predicted that both teachers would include more grammar-translation teaching in the first type of

class than in the second. He found, however, that while one teacher seemed to be influenced by the type of examination he was preparing his students for, the other explained grammar and employed translation no matter which type of examination he was dealing with.

- Shohamy et al. (1996) and Wall (1996) investigated the washback of high-stakes tests on teaching in state secondary schools. Shohamy et al. compared the effects of two tests in Israel, a test of English as a Foreign Language (EFL) and a test of Arabic as a Second Language (ASL), both when they were introduced and after they had been in place for some time. They found that while the washback of the EFL test increased over the years, the washback of the ASL test decreased ‘to the point where it has no effect: no special teaching activities are introduced in preparation for the test, no special time is allotted, no new teaching materials have been developed, awareness of the test is minimal ... ’ (Shohamy et al., 1996). The researchers attributed these changes to a variety of factors, including the status of the two languages within the country, the purposes of the tests, the test formats that were used and the skills that were tested.
- Wall (1996) reported on the washback of a major EFL test on secondary school teaching in Sri Lanka . Her study involved repeated observations at many schools over a two-year period, and in-depth interviews with the teachers whose classes were observed. She found that the test in question influenced the content of the classes in both positive and negative ways, but it had no influence on the methods the teachers used to deliver this content. Wall presented a number of reasons for these findings, relating not only to the test itself but also to other factors in the educational and social setting. These included a lack of understanding on many teachers’ part of the test construct, a lack of official feedback to teachers about their students’ test performance, and inadequate teacher support systems. Wall’s analysis was underpinned by insights from innovation theory (e.g. Fullan and Stiegelbauer, 1991).

Critical issues and topics :

There are a number of issues which have to be solved :

- The first has to do with the difficulty of separating out the influence of tests from the effects of other variables at work in the educational context.
- The second issue has to do with the dilemma many teachers face when they are preparing their students for a test that has important consequences. Most teachers work in a system with a student and a test is an individual and sometimes emotional one.
- The third issue has to do with the responsibilities of test developers with regard to the washback of their tests and/or any impact that extends beyond the classroom, into the educational system or even greater society.

Spratt (2005: 29) presents a long list of factors which have been shown to influence a test's washback. These include

- teacher-related factors (beliefs, attitudes, education and training),
- resourcing (with a focus on teacher-made and commercial materials),
- the conditions at the school where teaching and test preparation is taking place
- and the attributes of the test in question (e.g. its proximity, its purpose, the status of the language it tests, the formats it employs, the weighting of the different sections and how familiar the test is to teachers).

Spratt (2005: 24) calls 'a tension between pedagogical and ethical decisions'. The most effective solution to this problem is for test designers to 'sample widely and unpredictably' (Hughes, 2003: 54), in order to encourage teachers to teach all the points in the syllabus.

Current contributions and research:

Spratt (2005) lists the types of teacher beliefs that have been seen as important mediators of washback. These includes beliefs about Wall

- the reliability and fairness of the exam

- what constitutes effective teaching methods
- how much the exam contravenes their current teaching practices
- the stakes and usefulness of the exam
- their teaching philosophy
- the relationship between the exam and the textbooks, and
- their students' beliefs.

Only recently, have attempts been made to systematically study the nature of teachers'

beliefs and the influence that these can have on their classroom behaviour. Huang (2009) used insights from social psychology, in particular the **Theory of Planned Behaviour (Ajzen, 2006)**, to analyse data she had obtained through teacher diaries, observations, and in-depth interviews. Her purpose was to

- distinguish between teachers' behavioural beliefs (what they believed about the four types of teaching behaviour that a new high-stakes test was meant to encourage),
- their normative beliefs (what they believed important people around them – district inspectors, their head teacher, their peers, their students and the students' parents – expected of them with regard to the behaviours)
- and their control beliefs (what they believed was achievable given the possibilities and the constraints of the situation they were working in).

Another relatively new focus of study has to do with the washback of tests on learners, as seen through the eyes of learners. Washback studies have often been concerned with how tests influence teachers, perhaps because of the central role that teachers play in the classroom but also because of the practical difficulties of investigating student attitudes and behaviour.

Notable exceptions include Gosa (2004) and Tsagari (2009), who looked at students' reactions to tests as recorded in their diaries, and Watanabe (2001), who used interviews to explore the ways in which tests motivated