# THE WASHBACK EFFECT OF LANGUAGE TESTS
## JAMES DEAN BROWN
### *University of Hawai'i*

### *Introduction*

Numerous definitions of the concept of *washback* have been offered in the language teaching literature. For instance, Shohamy, Donitsa-Schmidt, and Ferman (1996, p. 298) define washback very simply as "the connections between testing and learning." Gates (1995, p. 101) defines it as "the influence of testing on teaching and learning." Shohamy (1992, p. 513) went further when she defined washback as "the utilization of external language tests to affect and drive foreign language learning...this phenomenon is the result of the strong authority of external testing and the major impact it has on the lives of test takers." Messick (1996, p. 241) provided an even more elaborate definition when he wrote, "washback, a concept prominent in applied linguistics, refers to 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."

One source of confusion about washback is that several different terms are used to refer to the connections between testing and learning. In the general educational literature the concept is referred to as *backwash*. Elsewhere the concept has been referred to as *test impact, measurement-driven instruction, curriculum alignment*, and *test feedback*. More humorously, at a meeting at Educational Testing Service, Dan Douglas once referred to the concept as the *bogwash* effect, a variant that I hope will endure.

A number of authors have linked washback to test validity. As Alderson and Wall (1993a, p. 116) point out, "some writers have even gone so far as to suggest that a test's validity should be measured by the degree to which it has had a beneficial influence on teaching." Messick (1996, p. 241) discusses "the concept of wasback as an instance of the consequential aspect of construct validity, linking positive washback to so-called authentic and direct assessments and, more basically, to the need to minimize construct underrepresentation and construct-irrelevant difficulty in the test." Morrow (1986) refers to something he calls *washback validity*, the relationship between a test and the related curriculum. Frederiksen and Collins (1989) discuss a similar concept but refer to it as *systemic validity*. Weir (1990, p. 27) suggests that communicative language testing could have a strong washback effect on communicative language teaching and, in fact, that such

a washback effect would be directly linked to the construct validity of the tests.

Because the first definition at the top of this paper (from Shohamy et al, 1996) is both adequate and parsimonious, it is very attractive. However, that definition does not explicitly include the link between washback and validity. Hence the working definition of washback that I will use in this paper is a slightly expanded version of the one provided by Shohamy et al (1996): the connections between language testing and learning, and the consequences of those connections.

I will continue to explore the concept of backwash by addressing a number of questions: Does washback exist? What factors affect the impact of washback? What are the negative aspects of washback? How can we promote positive washback? What directions might future research on washback effect take?

### *Does Washback Effect Exist?*

Quite reasonably, one of the main points of the article by Alderson and Wall (1993a) is that insufficient evidence was found in the literature at that time for the existence of washback. Indeed, their title asked that question directly: Does washback exist? They point to a great deal of literature that makes assertions about washback, but little actual empirical research into the existence and nature of washback. Similarly, Watanabe (1996b, p. 208) reports having found over 500 assertions about the impact of university entrance examinations in Japan, but only 10 empirical studies.

Alderson and Wall (1993a) point to four studies that empirically addressed the issue of washback in the past. Westdorp (1982) in the Netherlands investigated "the validity of objections to the introduction of multiple-choice tests into the assessment of mother tongue and foreign language education" and found that complaints based on assumed washback effects were not justified. Hughes (1988) in Turkey claimed a positive washback effect for an English proficiency test screening students to English medium university in Istanbul. Khaniya (1990) created a beneficial washback effect in Nepal and studied it. Alderson and Wall (1993a) note that the Westdorp (1982) report shows very little washback effect, while the other two studies are incomplete or inadequate in some way and lack investigations into "what actually changed in class" (p. 126). Another study they cite is Wall and Alderson (1993), which investigated the classroom impact of changing English examinations in Sri Lanka. That study found very little classroom impact. They suggest that this outcome may have resulted from teachers not understanding what is appropriate for preparing students for examinations. They also suggest that the examination "does not determine how teachers teacher, however much it might influence *what* they teach." (Alderson & Wall, 1993a, p. 127).

More recently, Alderson and Hamp-Lyons (1996) conducted a study in the United States of TOEFL preparation classes. They found that such classes are substantially different from non-TOEFL classes. TOEFL classes had more test-taking, more teacher talking time, less turn taking, less time spent on pair work, more references to TOEFL, more metalanguage, more routinized talk, and less laughter. They found that the picture was more complex when the teachers TOEFL and non-TOEFL classes were analyzed separately and that the ideas expressed about washback in the literature are too simplistic.

Shohamy, Donitsa-Schmidt, and Ferman (1996) investigated washback in Israel by examining the effects of an Arabic as a second language (ASL) examination for grades 7-9 and an EFL oral test for grade 12. They found the following effects for the ASL (p. 301):

1. Teachers stopped teaching new material and turned to reviewing material
2. Teachers replaced class textbooks with worksheets that were identical to previous years' tests
3. The activities were all "testlike"
4. Review sessions were added to regular class hours
5. The atmosphere in the class was tense
6. Teachers and students were highly motivated to master the material.
7. When the test was over the above activities stopped

They found quite different effects for the EFL oral test (p. 301):

1. Experienced teachers spent more time on teaching oral language
2. Experienced teachers used only activities identical to the ones on the test
3. Novice teachers tried out additional oral language activities

Watanabe (1992) examined washback in Japan and found that, if it exists, it broadens the range of strategies that students will use, an effect that persists one year later. Later, Watanabe (1996a) investigated washback in the classrooms of two *yobikoi* teachers in Japan. His goal was to study the use of exam induced translation in class. He concluded that translation on exams affects some teachers and not others depending on personal beliefs, educational background, and past learning experiences (p. 330). He also felt that claims of washback may be exaggerated and somewhat inconsistent. Watanabe (1996b) cited a number of other related papers Ariyoshi and Senba (1983), Fujita (1992), Saito et al (1984) with bearing on the issue of washback. Unfortunately, I have been unable to obtain those papers at the time of this writing.

Wall (1996) spent four years in a Sri Lankan EFL project evaluating a new national examination. Similar to what was reported in Wall and Alderson (1993), Wall (1996) reports:

The main findings... were that the examination had had considerable impact on the content of English lessons and on the way teachers design their classroom tests (some of this was positive and some negative), but it had had little to no impact on the methodology they used in the classroom or on the way they marked their pupils' test performance.

Wall found the language testing advice of Hughes (1989), Shohamy (1992), and Bailey (1996) and the general education recommendations of Heyneman and Ransom (1990) and Kellaghan and Greaney (1992) helpful in improving curriculum (p. 350-351)

All in all, the empirical studies to date indicate that the washback effect does exist in various forms in various places, but also that the issue is not a straightforward one that conforms neatly to the popular notions of the effects of examinations on language learning.

### What Factors Affect the Impact of Washback?

Given the complexity of the issues involved in washback, there must be a number of factors that impact on it. Gates (1995, p. 101) outlines two ways in which washback may vary: it can range from positive to negative washback and from strong to weak washback. The question raised in this section is what factors cause tests to have positive or negative, weak or strong washback effects? Gates suggests (pp. 102-103) that the following factors affect the impact of washback:

1. Prestige
2. Accuracy
3. Transparency
4. Utility
5. Monopoly
6. Anxiety
7. Practicality

Alderson and Hamp-Lyons (1996, p. 296) argue that the amount and type of washback depend on the extent to which:

1. The test has status (and level of stakes)
2. The test is counter to current teaching practices
3. Teachers and textbook writers think about appropriate methods of test preparation
4. Teachers and textbook writers are willing and able to innovate

Shohamy et al (1996, pp. 299-300, 314-315) say that the degree of impact is influenced by:

1. Status of the subject-matter (language)
2. Low vs. high stakes
3. Nature of the test (purpose)
4. Format of the test (more anxiety from oral vs. written test; novel vs. familiar format; etc.)
5. Use to which the scores will be put
6. Skills being tested: "It may very well be that in multiskilled tests, test-takers may feel more confident since they can compensate for a lack of proficiency in one skill by high proficiency in the other"

Shohamy et al. (1996, p. 303) also suggest ways of judging the impact of tests on curriculum:

1. Classroom activities and time allotment
2. The extent to which the test has generated new teaching materials
3. The degree to which students and parents are aware of the existence and content of the test
4. Perceived effects of test results
5. The extent to which the test has changed the prestige and position of the areas tested
6. Perception of test quality and importance
7. Impact of test on promoting learning
8. How the various language inspectors view the role, status, and impact of the test

Table 1 summarizes the factors that affect the impact of washback. Surprisingly, very little overlap exists between the lists provided by the authors cited above. Only two of the 24 items listed in the table were mentioned in two articles. Notice, in Table 1, that I have organized the factors into four categories: prestige factors, test factors, people factors, and curriculum factors.

### What Are the Negative Aspects of Washback?

Many language educators believe that tests have negative washback effects on the learning and teaching of languages. These beliefs are often based on assumptions like those pointed out by Watanabe as underlying the washback effects of university of entrance examinations in Japan (1996a, p. 319):

1. A substantial number of the questions in university entrance examination require grammar and translation skills (GT)

Table 1
*Factors That Affect the Impact of Washback*

---

***Prestige Factors***
  1. Prestige or status of the test (Gates, 1995; Alderson & Hamp-Lyons, 1996)
  2. Status of the subject-matter of the test (Shohamy et al, 1996)
  3. Perception of quality and importance of the test (Shohamy et al, 1996)
  4. Degree to which a test has a monopoly on assessment (Gates, 1995)
  5. Level of stakes (Alderson & Hamp-Lyons, 1996; Shohamy et al, 1996)
  6. Use to which the scores will be put (Shohamy et al, 1996)

***Test Factors***
  1. Nature of the test (purpose) (Shohamy et al, 1996)
  2. Format of the test (Shohamy et al, 1996)
  3. Skills being tested (Shohamy et al, 1996)
  4. Accuracy the scores (Gates, 1995)
  5. Utility of the test and its results (Gates, 1995)
  6. Practicality of the test (Gates, 1995)

***People Factors***
  1. The degree to which students and parents are aware of the existence and content of the test (Shohamy et al, 1996)
  2. How students and parents perceive effects of test results (Shohamy et al, 1996)
  3. Student anxiety (Gates, 1995)
  4. How the various language inspectors view the role, status, and impact of the test (Shohamy et al, 1996)
  5. The extent to which the test has changed the prestige and position of the areas tested (Shohamy et al, 1996)
  6. Transparency of the information provided by the test to teachers, students, parents, etc. (Gates, 1995)

***Curriculum Factors***
  1. Match of test to current teaching practices (Alderson & Hamp-Lyons, 1996)
  2. Effect of test on promoting learning (Shohamy et al, 1996)
  3. Types of classroom activities and time allotted for each (Shohamy et al, 1996)
  4. The extent to which the test has generated new teaching materials (Shohamy et al, 1996)
  5. The extent to which teachers and textbook writers think about appropriate methods of test preparation (Alderson & Hamp-Lyons, 1996)
  6. The extent to which teachers and textbook writers are willing and able to innovate (Alderson & Hamp-Lyons, 1996)

2. This is why so much GT is employed in the classroom
3. Unless the content of these exams changes, nothing will change in the teaching of EFL
4. Corollary: "…if the exam begins to use other types of test questions, then teachers will use methods other than GT." (p. 319)

Such thinking may often underlie beliefs in the negative washback effects. But more generally, what are the negative washback effects that language educators perceive?

Alderson and Hamp-Lyons (1996) in their study about TOEFL feel that the following were negative results of washback (p. 280):

1. Unnatural teaching
2. Students being taught inappropriate language-learning and language-using strategies
3. Students being taught "TOEFLese"
4. Courses that raise TOEFL scores without providing students with the English they will need in language interaction or in the college or university courses they are entering
5. Students taking TOEFL courses instead of "real" English courses

They also found four main themes in the washback literature (p. 281), which could be construed as negative washback effects:

1. Narrowing the curriculum
2. Lost instructional time
3. Reduced emphasis on skills that require complex thinking or problem-solving
4. Test score 'pollution', or increases in test scores without an accompanying rise in ability in the construct being tested

Bailey (1996, p. 264-265) discusses ten choices that students might make due to washback (those with an asterisk might be construed as negative choices):

1. *Practicing items similar in format to those on the test
2. *Studying vocabulary and grammar rules
3. Participating in interactive language practice (e.g., target language conversations)
4. Reading widely in the target language
5. Listening to non-interactive language (radio, television, etc.)
6. *Applying test-taking strategies
7. *Enrolling in tests-preparation courses
8. Requesting guidance in their studying and feedback on their performance
9. *Enrolling in, requesting or demanding additional (unscheduled) test-preparation classes or tutorials (in addition to or in lieu of other language classes)

10. *Skipping language classes to study for the test

Shohamy et al (1996), in investigating the washback effect in Israel, found the following negative effects for a national Arabic as a second language test for grades 7-9 before it was administered (p. 301):

1. Teachers stopped teaching new material and turned to reviewing material
2. Teachers replaced class textbooks with worksheets that were identical to previous years' tests
3. The activities were all "testlike"
4. Review sessions were added to regular class hours
5. The atmosphere in the class was tense

However, according to the authors, once the test had been administered, the above activities stopped.

Table 2 summarizes the negative effects of washback. Again, surprisingly, very little overlap exists between the lists given above. Only two of the 16 factors listed in the table were mentioned in two articles. Notice, in Table 2, that I have organized the factors affecting washback into four categories: teaching factors, course content factors, course characteristic factors, and class time factors.

### How Can We Promote Positive Washback?

As pointed out above, Shohamy et al (1996), in their investigation of the washback effect of an EFL oral test in Israel, find the following positive washback effects (p. 301):

1. Experienced teachers spent more time on teaching oral language
2. Experienced teachers used only activities identical to the ones on the test
3. Novice teachers tried out additional oral language activities

Clearly, tests can and sometimes do produce positive washback effects.

Gates (1995, p. 101) suggests that "teachers might reasonably want to determine the type of washback that flows from a given test." They should be interested so they can limit negative washback effects and promote positive washback. A number of authors have addressed this latter issue.

Hughes (1989, p. 1) provides an entire chapter discussing the following ways of promoting beneficial backwash:

1. Test the abilities whose development you want to encourage
2. Sample widely and unpredictably
3. Use direct testing
4. Make testing criterion-referenced
5. Base achievement tests on objectives

Table 2

*Negative Aspects of Washback*

---

***Teaching Factors***
1. Teachers narrow the curriculum (Alderson & Hamp-Lyons, 1996)
2. Teachers stop teaching new material and turned to reviewing material (Shohamy et al, 1996)
3. Teachers replaced class textbooks with worksheets identical to previous years' tests (Shohamy et al, 1996)
4. Unnatural teaching (Alderson & Hamp-Lyons, 1996)

***Course Content Factors***
1. Students being taught "examination-ese" (Alderson & Hamp-Lyons, 1996)
2. Students Practicing "testlike" items similar in format to those on the test (Bailey, 1996; Shohamy et al, 1996)
3. Students applying test-taking strategies in class (Bailey, 1996)
4. Students studying vocabulary and grammar rules [to the exclusion of other aspects of language] (Bailey, 1996)

***Course Characteristic Factors***
1. Students being taught inappropriate language-learning and language-using strategies (Alderson & Hamp-Lyons, 1996)
2. Reduced emphasis on skills that require complex thinking or problem-solving (Alderson & Hamp-Lyons, 1996)
3. Courses that raise examination scores without providing students with the English they will need in language interaction or in the college or university courses they are entering; also called this test score 'pollution' (Alderson & Hamp-Lyons, 1996)
4. The tense atmosphere in the class (Shohamy et al, 1996)

***Class Time Factors***
1. Enrolling in, requesting or demanding additional (unscheduled) test-preparation classes or tutorials (in addition to or in lieu of other language classes) (Alderson & Hamp-Lyons, 1996; Bailey, 1996)
2. Review sessions added to regular class hours (Shohamy et al, 1996)
3. Skipping language classes to study for the test (Bailey, 1996)
4. Lost instructional time (Alderson & Hamp-Lyons, 1996)

---

6.  Ensure test is known and understood by students and teachers
7.  Where necessary provide assistance to teachers

Heyneman and Ransom (1990, p. 112) suggest three strategies for improving test content so as to create more positive washback effects:

1.  Use more open-ended items (as opposed to selected-response items like m-c)
2.  Test higher-level cognitive skills
3.  Authorities should provide feedback to teachers and others (teacher trainers, curriculum developers, inspectors, education officers, and head teachers) so meaningful change can be effected

Kellaghan and Greaney (1992, p. 3), in reviewing World Bank research in 14 countries in Africa, suggest the following to lessen negative washback on classroom teaching (as summarized in Wall, 1996, p. 337):

1.  Examinations should reflect the full curriculum, not merely a limited aspect of it.
2.  Higher-order cognitive skills should be assessed to ensure they are taught.
3.  Skills to be tested should not be limited to academic areas; should also relate to out-of-school tasks.
4.  A variety of examinations formats should be used, including written, oral, aural, and practical.
5.  In evaluating published examination results and national rankings, account should be taken of factors other than teaching effort.
6.  Detailed, timely feedback should be provided to schools on levels of pupis performance and areas of difficulty in public examinations.
7.  Predictive validity studies of public examinations should be conducted. (This is to see whether selected exams are fulfilling their purpose).
8.  The professional competence of examination authorities needs improvement, especially in test design.
9.  Each examination board should have a research capacity. (This is to investigate, among other things, the impact of examinations on teaching.)
10. Examination authorities should work closely with curriculum organizations and with educational administrators.
11. Regional professional networks should be developed to initiate exchange programmes and to share common interests and concerns.

Bailey (1996, pp. 268-269) suggests that we could promote beneficial washback by incorporating the following into our tests:

1.  Language learning goals
2.  Authenticity

3. Learner autonomy and self-assessment
4. Detailed score reporting

She (p. 275) also lists other criteria likely to promote beneficial washback:

1. Test-takers, teachers, administrators, and curriculum designers should understand the purpose of the test
2. Results must be believable to test takers and score users
3. Test takers must find the results credible and fair
4. Test should measure what the programs intend to teach

Bailey concludes (p. 275) that a test will promote beneficial washback to the extent that:

1. It is based on sound theoretical principles
2. It uses authentic tasks and texts
3. Test takers buy into the assessment process

Wall (1996), in reviewing the literature, lists the desirable characteristics in language testing as being the following (p. 334):

1. Direct testing
2. Criterion-referencing
3. Authentic texts
4. Tasks

Drawing on other authors, Wall further suggests improving the washback effect by doing the following (pp. 334-335):

1. Teachers and students should understand the tests for which they are preparing (Hughes, 1989)
2. Teachers should receive help so they understand the tests (Hughes, 1989)
3. Schools should receive feedback from testers (Shohamy, 1992) and
4. Teachers and administrators should be involved in different phases of the testing process because they are the people who will have to make changes (Shohamy, 1992)

Table 3 summarizes the ways suggested in the literature to promote positive washback effects. Again, surprisingly little overlap exists between the lists given above. Only five of the 28 items listed in the table were mentioned in two articles. Notice, in Table 3, that I have organized the factors that promote positive washback into four categories: test design factors, test content, logistic factors, as well as interpretation and analysis factors.

Table 3
*Promoting Beneficial Backwash*

---

*Test Design Factors*
1.  Sample widely and unpredictably (Hughes, 1989)
2.  Design tests to be criterion-referenced (Hughes, 1989; Wall, 1996)
3.  Design the test to measure what the programs intend to teach (Bailey, 1996)
4.  Base the test on sound theoretical principles (Bailey, 1996)
5.  Base achievement tests on objectives (Hughes, 1989)
6.  Use direct testing (Hughes, 1989; Wall, 1996)
7.  Foster learner autonomy and self-assessment (Bailey, 1996)

*Test Content Factors*
1.  Test the abilities whose development you want to encourage (Hughes, 1989)
2.  Use more open-ended items (not selected-response items like m-c) (Heyneman & Ransom, 1990)
3.  Make examinations reflect the full curriculum, not a limited part (Kellaghan & Greaney, 1992)
4.  Assess higher-order cognitive skills to ensure they are taught (Heyneman & Ransom, 1990; Kellaghan & Greaney, 1992)
5.  Use a variety of examination formats, including written, oral, aural, and practical (Kellaghan & Greaney, 1992)
6.  Do not limit skills to be tested to academic areas (should also relate to out-of-school tasks) (Kellaghan & Greaney, 1992)
7.  Use authentic tasks and texts (Bailey, 1996; Wall, 1996)

*Logistic Factors*
1.  Insure that test-takers, teachers, administrators, curriculum designers understand the purpose of the test (Bailey, 1996; Hughes, 1989)
2.  Make sure language learning goals are clear (Bailey, 1996)
3.  Where necessary provide assistance to teachers to help them understand the tests (Hughes, 1989)
4.  Provide feedback to teachers and others so meaningful change can be effected (Heyneman & Ransom, 1990; Shohamy, 1992)
5.  Provide detailed and timely feedback to schools on levels of pupils' performance and areas of difficulty in public examinations (Kellaghan & Greaney, 1992)
6.  Make sure teachers and administrators are involved in different phases of the testing process because they are the people who will have to make changes (Shohamy, 1992)
7.  Provide detailed score reporting (Bailey, 1996)

*Interpretation and Analysis Factors*
1.  Make sure results are believable, credible, and fair to test takers and score users (Bailey, 1996)
2.  Consider factors other than teaching effort in evaluating published examination results and national rankings (Kellaghan & Greaney, 1992)
3.  Conduct predictive validity studies of public examinations (This is to see whether selected exams are fulfilling their purpose) (Kellaghan & Greaney, 1992)
4.  Improve the professional competence of examination authorities, especially in test design (Kellaghan & Greaney, 1992)
5.  Insure that each examination board has a research capacity (In order to investigate, among other things, the impact of examinations on teaching) (Kellaghan & Greaney, 1992)
6.  Have examination authorities work closely with curriculum organizations and with educational administrators (Kellaghan & Greaney, 1992)
7.  Develop regional professional networks to initiate exchange programs and to share common interests and concerns (Kellaghan & Greaney, 1992)

---

### What Directions Might Future Research on Washback Effect Take?

Earlier, I showed how the literature: supports the existence of washback effects in various situations, but also reveals that the issues involved are far from simple. Many authors have listed the factors they think affect the impact of washback, factors they think are the negative aspects of washback, and factors they think promote positive washback. However, as I was summarizing these three aspects of the literature in Tables 1-3, I realized that very little overlap exists among the lists of different authors. In other words, little agreement was found about what factors affect washback, what the negative aspects are, and what we can do to promote positive washback. Clearly, much more research is needed in this important area of language testing, especially research that can clarify the above three issues.

Alderson and Wall (1993, pp. 120-121) suggest 15 hypotheses that should be investigated in this regard:

1. A test will influence teaching
2. A test will influence learning
3. A test will influence *what* teachers teach
4. A test will influence *how* teachers teach
5. A test will influence *what* learners learn
6. A test will influence *how* learners learn
7. A test will influence the *rate* and *sequence* of teaching
8. A test will influence the *rate* and *sequence* of learning
9. A test will influence the *degree* and *depth* of teaching
10. A test will influence the *degree* and *depth* of learning
11. A test will influence the attitudes to the content, method, etc. of teaching and learning
12. Tests that have important consequences will have washback
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 teachers, but *not* for others

Watanabe (1996b) suggests five research questions:

1. Does washback exist?
2. What evidence enables us to say washback exists or does not exist?
3. If washback exists, what is its nature (i.e., positive or negative)?
4. If washback does not exist, why not?
5. If washback exists, under what conditions?

Bailey (1996, pp. 276-277) offers questions for research within a language program:

1. Do the participants understand the purpose(s) of the test and the intended use(s) of the results?
2. Are the results provided in a clear, informative and timely fashion?
3. Are the results perceived as believable and fair by the participants?
4. Does the test measure what the program intends to teach?
5. Is the test based on clearly articulated goals and objectives?
6. Is the test based on sound theoretical principles which have current credibility in the field?
7. Does the test utilize authentic texts and authentic tasks?
8. Are the participants invested in the assessments processes?

Shohamy et al (1996, p. 298) raised questions of their own:

1. Is introducing changes through tests effective?
2. Can the introduction of tests *per se* cause real improvement in learning and teaching?
3. How are test results used by teachers, students and administrators.

Table 4 summarizes some questions that future research on the washback effect might profitably address. Again, surprisingly little overlap appears to exist between the lists given by various authors. Only one of the 31 questions listed in the table were mentioned in two articles. Notice, in Table 4, that I have organized the questions for future research into three categories: general questions, detailed questions, and program related questions.

### Conclusion

In this article, I reviewed several definitions of washback. Then, I set out to answer a number of questions and found the following:

1. *Does washback exist?* The answer was a qualified *yes*. The literature supports the notion that washback exists in various places in various ways. Clearly, the issue is a complex one that warrants considerably more research.
2. *What factors affect the impact of washback?* I identified a total of 24 factors in the literature that have an impact on washback. These 24 included prestige factors, test factors, people factors, and curriculum factors.
3. *What are the negative aspects of washback?* I found a total of 16 factors in the literature that seem to be negative aspects of washback. These 16 included teaching factors, course content factors, course characteristic factors, and class time factors.

Table 4
*Directions for Future Research on Washback*

---

### General Questions

1.  Does washback exist? (Alderson & Wall, 1993; Watanabe, 1996b)
2.  What evidence enables us to say washback exists or does not exist? (Watanabe, 1996b)
3.  If washback exists, what is its nature (i.e., positive or negative)? (Watanabe, 1996b)
4.  If washback does not exist, why not? (Watanabe, 1996b)
5.  If washback exists, under what conditions? (Watanabe, 1996b)
6.  Is introducing changes through tests effective? (Shohamy et al, 1996)
7.  Can the introduction of tests *per se* cause real improvement in learning and teaching? (Shohamy et al, 1996)
8.  How are test results used by teachers, students and administrators? (Shohamy et al, 1996)

### Detailed Questions

1.  Will test influence teaching (Anderson & Wall, 1993)
2.  Will test influence learning (Anderson & Wall, 1993)
3.  Will test influence *what* teachers teach (Anderson & Wall, 1993)
4.  Will test influence *how* teachers teach (Anderson & Wall, 1993)
5.  Will test influence *what* learners learn (Anderson & Wall, 1993)
6.  Will test influence *how* learners learn (Anderson & Wall, 1993)
7.  Will test influence the *rate* and *sequence* of teaching (Anderson & Wall, 1993)
8.  Will test influence the *rate* and *sequence* of learning (Anderson & Wall, 1993)
9.  Will test influence the *degree* and *depth* of teaching (Anderson & Wall, 1993)
10. Will test influence the *degree* and *depth* of learning (Anderson & Wall, 1993)
11. Will test influence the attitudes to the content, method, etc. of teaching and learning (Anderson & Wall, 1993)
12. Will tests that have important consequences have more washback (Anderson & Wall, 1993)
13. Will tests that do not have important consequences have no washback (Anderson & Wall, 1993)
14. Will tests have washback on *all* learners and teachers (Anderson & Wall, 1993)
15. Will tests have washback effects for *some* learners and teachers, but *not* for others (Anderson & Wall, 1993)

### Program-Related Questions

1.  Do the participants understand the purpose(s) of the test and the intended use(s) of the results? (Bailey, 1996)
2.  Are the results provided in a clear, informative and timely fashion? (Bailey, 1996)
3.  Are the results perceived as believable and fair by the participants? (Bailey, 1996)
4.  Does the test measure what the program intends to teach? (Bailey, 1996)
5.  Is the test based on clearly articulated goals and objectives? (Bailey, 1996)
6.  Is the test based on sound theoretical principles which have current credibility in the field? (Bailey, 1996)
7.  Does the test utilize authentic texts and authentic tasks? (Bailey, 1996)
8.  Are the participants invested in the assessments processes? (Bailey, 1996)

---

4. *How can we promote positive washback?* I unearthed a total of 28 factors in the literature that promote positive washback. These 28 included test design factors, test content factors, logistic factors, and interpretation and analysis factors.

5. *What directions might future research on washback effect take?* I also compiled a list of 31 general, detailed, and program related research questions that I found in the literature research questions that need to be answered.

Answers to the five questions addressed in this article are important if we are to responsibly use tests for making important decisions about our students. As Watanabe (1996, p. 332) put it, "a large amount of time, money and energy is spent on entrance exams every year at individual, school and national levels. In order to make the best use of such an investment, we need to be empirical, rational and well informed."

## REFERENCES

Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing, 13*, 280-297.

Alderson, J. C., & Wall, D. (1993a). Does washback exist? *Applied Linguistics, 14*, pp. 115-129.

Alderson, J. C., & Wall, D. (1993b). Examining washback: The Sri Lankan impact study. *Language Testing, 10*, 41-69.

Ariyoshi, H., & Senba, K. (1983). Daigaku nyushi junbi kyoiku ni kansuru kenkyu (A study on preparatory teaching for the university entrance examinations). *Fukuoka Kyoiku Daigaku Kiyo, 33*, 1-21.

Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing, 13*, 257-279.

Berwick, R., & Ross, S. (1989). Motivation and matriculation: Are students still alive after exam hell? *JALT Journal, 11*(2), 193-210.

Buck, G. (1988). Testing listening comprehension in Japanese university entrance examinations. *JALT Journal, 10*, 15-42.

Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Education researcher, 18*(9), 27-32.

Fujita, T. (1992). Readiness model of optimal input: A comparison between Japanese high school and non-high school students of English. *Sophia Linguistica, 31*, 122-143.

Gates, S. (1995). Exploiting washback from standardized tests. In J. D. Brown & S. O. Yamashita (Eds.), *Language testing in Japan* (pp. 101-106). Tokyo: Japanese Association for Language Teaching.

Green, D. Z. (1985). Developing measures of communicative proficiency: A test for French immersion students in grades 9 and 10. In P. C. Hauptman, R. LeBlanc, & M. B. Wesche (Eds.), *Second language performance testing* (pp. 215-227). Ottawa, Canada: University of Ottawa Press.

Hart, D., Lapkin, S., & Swain, M. (1987). Communicative language tests: Perks and perils. *Evaluation and Research in Education, 1,* 83-93.

Heyneman, S. P., & Ransom, A. W. (1990). Using examinations and testing to improve educational quality. *Educational Policy,* 177-192.

Hughes, A. (1988). Introducing a needs-based test of English language proficiency into an English-medium university in Turkey. In A. Hughes (Ed.), *Testing English for University Study.* ELT Document #127. Modern English Publications.

Hughes, A. (1989). *Testing for language teachers.* Cambridge: Cambridge University Press.

Jones, R. L. (1985). Second language performance testing: An overview. In P. C. Hauptman, R. LeBlanc, & M. B. Wesche (Eds.), *Second language performance testing* (pp. 15-24). Ottawa: University of Ottawa Press.

Khaniya, T. R. (1990). The washback effect of a textbook-based test. *Edinburgh Working Papers in Applied Linguistics, 1,* 48-58.

Kellaghan, T., & Greaney, V. (1992). *Using examinations to improve education: A study of fourteen African countries.* Washington, DC: The World Bank.

Law, G. (1994). College entrance exams and team teaching in high school English classrooms. In M. Wada & T. Cominos (Eds.), *Studies in team teaching* (pp. 90-102). Tokyo: Kenkyusha.

Law, G. (1995). Ideologies of English language education in Japan. *JALT Journal, 17*(2), 213-224.

Mehrens, W. A., & Kaminsky, J. (1989). Methods for improving standardized test scores: Fruitful, fruitless, or fraudulent? *Educational Measurement: Issues and Practice, 8,* 14-22.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18,* 5-11.

Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13,* 241-256.

Miller, M. D., & Legg, S. M. (1993). Alternative assessment in a high-stakes environment. *Educational Measurement: Issues and Practice, 12,* 9-15.

Morrow, K. (1986). The evaluation of tests of communicative performance. In M. Portal (Ed.), *Innovations in Language Testing.* London: NFER/Nelson.

Morrow, K. (1991). Evaluating communicative tests. In S. Anivan (Ed.), *Current developments in language testing* (pp. 111-118). Singapore: Regional Language Centre.

Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research, 62*(3), 229-258.

Powers, D. E. (1993). Coaching for the SAT: A summary of the summaries and an update. *Educational Measurement: Issues and Practice, 12*, 24-30.

Reader, I. (1986). Language teaching in Britain and Japan: A personal view. *JALT Journal, 7*(2), 113-136.

Rohlen, T. *Japan's high schools*. Berkley, CA: University of California Press.

Saito, T., Ariata, S., & Nasu, I. (1984). Taki-sentaku tesuto ga igaku kyoiku ni oyobosu eikyo (Effects of multiple-choice questions on medical education). *Nihon igaku kyoiku shinko zaidan kenkyu jose ni yoru kenkyu hokoku sho.*

Shohamy, E. (1992). Beyond performance testing: A diagnostic feedback testing model for assessing foreign language learning. *Modern Language Journal, 76*(4), 513-521.

Shohamy, E. (1995). Performance assessment in language testing. *Annual Review of Applied Linguistics, 15*, 188-211.

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

Swain, M. (1984). Large-scale communicative testing: A case study. In S. J. Savignon & M. Berns (Eds.), *Initiatives in communicative language teaching* (pp. 185-201). Reading, MA: Addison-Wesley.

Swain, M. (1985). Large-scale communicative testing. In Y. P. Lee, C. Y. Y. Fok, R. Lord, & G. Low (Eds.), *New directions in language testing*. Hong Kong: Pergamon Press.

Wall, D. (1996). Introducing new tests into traditional systems: Insights from general education and from innovation theory. *Language Testing, 13*, 234-354.

Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing, 10*, 41-69.

Wall, D., & Alderson, J. C. (1996). Examining washback: The Sri Lankan impact study (pp. 194-221). In A. Cumming & R. Berwick (Eds.), *Validation in language testing*. Clevedon, UK: Multilingual Matters.

Watanabe, Y. (1992). Washback effects of college entrance examinations on language learning strategies. *JACET Bulletin, 23*, 175-194.

Watanabe, Y. (1996a). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing, 13*, 318-333.

Watanabe, Y. (1996b). Investigating washback in Japanese EFL classrooms: Problems and methodology. *Australian Review of Applied Linguistics, 13*, 208-239.

Watson-Gegeo, K. A. (1988). Ethnography in ESL: Defining the essentials. *TESOL Quarterly, 22*(4), 575-592.

Weir, C. J. (1990). *Communicative language testing*. Hemel Hempstead, UK: Prentice Hall.

Wesdorp, H. (1982). *Backwash effects of language testing in primary and seconday education*. Unpublished ms. Stichting Centrum voor onderwijsonderzoek van de Universiteit van Amserdam, The Netherlands.

James Dean Brown
Department of ESL
1890 East-West Road
University of Hawai'i
Honolulu, Hawai'i 96822

e-mail: brownj@hawaii.edu