

A critical review of five language washback studies from 1995-2007: Methodological considerations

by Yi-Ching Pan

National Pingtung Institute of Commerce, Taiwan

The University of Melbourne, Australia

Abstract

This paper focuses on five different washback studies during the last decade. Starting with a brief discussion of Messick's 1996 seminal work on the consequential aspect of construct validity and its relevance to washback, we will explore the contributions of Shohamy et al., Alderson & Hamp-Lyons, Chen, Green and Shi to the notion of washback and test validity. Each study is evaluated in terms of its contribution to our current understanding of washback. Finally, suggestions are made for future washback studies.

Keywords: washback, examination consequences, test validity, construct validity, consequential validity

In his acclaimed paper regarding validity, Messick (1989) developed the concept of consequential validity, changing our notions about score interpretation and test use. The concept of washback in test validity research is often associated with Messick's concept of consequential validity. Messick (1996) viewed washback as an "instance of the consequential aspect of construct validity" (p. 242), which covers elements of test use, the impact of testing on test-takers and educators, the interpretation of results by decision-makers, and any possible misuses, abuses, and unintentional effects of tests (Messick, 1989). Washback has become a focal point of validity research in that Messick (1996) contends that it is a component of the consequential aspect of construct validity and as such must be factored into any evaluation of validity. The effects of tests on teachers, students, institutions, and society are accordingly considered one type of validity evidence. Many other researchers (Bachman, 2005; Cronbach, 1988; Kane, 1992; McNamara & Roever, 2006; Shohamy, 2001) have also stressed the importance of justifying test use and investigating its consequences.

"Washback has become a focal point of validity research . . ."

This paper focuses on five language washback studies published between 1995 and 2007. The first section summarizes how these studies have explored washback. Each study is then evaluated in terms of the way it has enhanced our understanding of the scope and nature of washback. Finally, suggestions are made for future washback studies. These studies, summarized in Table 1, are discussed chronologically and examined because they offer methodological considerations as to how to gauge various aspects of washback from different stakeholders.

Table 1. Major Language Washback Studies Published in English from 1995-2007

Studies:	Shohamy, et al. (1996)	Alderson & Hamp-Lyons (1996)	Cheng (1999)	Green (2007)	Shih (2007)
Exams studied:	An Israeli ASL test & ESL test	TOEFL [®]	Old & New HKCEE	IELTS Writing Test	GEPT
Purposes:	To examine the impact of 2 national tests in and beyond classroom settings	To ascertain the influence of the TOEFL [®] on class teaching	To compare teachers' perceptions toward both exams	To examine how preparation classes impact score gains	To explore the effects of GEPT exit requirements on learning
Methodologies:	<ol style="list-style-type: none"> 1. Student questionnaires 2. Structured interviews with teachers and inspectors 3. Analysis of inspectorate bulletins 	<ol style="list-style-type: none"> 1. Interviews with teachers and students 2. Classroom observations 	<ol style="list-style-type: none"> 1. Teacher /student questionnaires 2. Structured interviews with teachers 3. Classroom observations 	<ol style="list-style-type: none"> 1. Two IELTS writing tests 2. Two questionnaires consisting of participant and process variables respectively 	<ol style="list-style-type: none"> 1. Interviews with department heads teachers, students, and family members 2. Classroom observations
Collected evidence:	<ol style="list-style-type: none"> 1. More positive washback found in ESL 2. More negative washback found in ASL 	<ol style="list-style-type: none"> 1. More occurrences of teacher talk, the use of meta language in non-TOEFL[®] classes 2. Fewer opportunities for pair work, laughter, and turn-taking in TOEFL[®] classes 	<ol style="list-style-type: none"> 1. An increased change in teaching content and activities 2. A lack of change in teaching methodologies 	An improvement in test scores for learners in test-preparation or academic-oriented classes, but those in the former progressed no more than those in the latter	<ol style="list-style-type: none"> 1. Small but varied aspects of washback found in students at both schools with and without exit requirements 2. External, intrinsic and test factors explain GEPT's minor impact on students' learning
Conclusions:	Washback changes over time because of factors including language status and test uses.	TOEFL [®] affects both what and how teachers teach, but the effect varies with teachers.	The change on teaching content rather than methodology was attributed to inadequate training and qualifications of secondary English teachers.	Test preparation classes have no apparent benefit to improve test scores.	The current washback theory didn't account for GEPT washback, so a new learning washback model has been developed.

Shohamy, et al. (1996) – ASL and EFL tests in Israel

Shohamy et al. (1996) examined the impact of national tests of Arabic as a Second Language (ASL) and English as a Foreign Language (EFL) in Israel. They explored different washback patterns among teachers, students, and inspectors in terms of how these tests influenced classroom activities, time allotment, teaching materials, perceptions of prestige, and the overall enhancement of learning. Regarding the EFL test, oral teaching activities were progressively introduced. As a consequence the amount of instruction time for oral activities increased, new courseware was brought in, awareness of the test increased, and the subject matter's status in the school substantially rose. In contrast, the ASL's impact in those areas declined to the point of insubstantiality. Nevertheless, the bureaucrats believed both tests had reached their objectives without any need for teacher training or curricular revision. The study concludes that washback changes with time because of factors such as language status and test uses, a finding which has since been corroborated by other researchers such as Stoneman (2006), who investigated how students prepared for a Hong Kong exit exam. The results of Stoneman's investigation showed that higher-status exams such as the International English Language Testing System (IELTS) motivated students to study more than lower-status tests such as Hong Kong's Graduating Students Language Proficiency Assessment (GSLAP).

Alderson & Hamp-Lyons (1996) – TOEFL® and non-TOEFL® preparation classes in North America

This study of the influence of the TOEFL on classroom teaching utilized interviews with 11 teachers in the U.S. and an unspecified number of students, as well as observations of 8 regular and 8 preparatory classes taught by the same two teachers. Results revealed that non-TOEFL® classes exhibited more student questioning and a greater degree of student-student and student-teacher interaction. TOEFL classes, on the other hand, showed fewer digressions and less laughter, and teachers tended to teach to the test. Moreover, Alderson & Hamp-Lyons claimed that the TOEFL influences both what and how teachers teach, but the effect varies in degree or in type among teachers.

Cheng (1999; 2004) – Old and new HKCEEs at secondary schools in Hong Kong

This study investigated the possible washback effects from the 1994 Revised Hong Kong Certificate of Education Exam in English (HKCEE) on teachers and students in Hong Kong secondary schools. Classroom observations of 12 high school teachers for 45 lessons, as well as questionnaires by 550 teachers and 1700 students, and interviews with an unspecified number of teachers reveal a range of attitudes and behavioral changes over the 1994-1995 period. The ostensible intention of the exam reform was to inspire integrated, task-based teaching.

Cheng, however, determined from the questionnaires that although most teachers felt positively about the revised exam that enabled students to use English more practically and authentically, no major changes emerged in terms of actual pedagogic practices, which are still content-based and teacher-centered. The content of what was taught now focuses more on listening and speaking in accordance with the revised exam. As Cheng stated, “the change of the HKCEE toward an integrated and task-based approach showed teachers the possibility of something new, but it did not automatically enable teachers to teach something new” (p. 164). Cheng’s study confirms Wall and Alderson’s (1993) previous findings: while classroom content may change because of a test, the way teachers instruct does not change to any significant degree. The changes noted by Cheng (2005, p. 235) were “superficial”.

Green (2007) – IELTS preparation, pre-session EAP, and combination thereof in Britain

This study investigated whether test preparation classes were advantageous in assisting students trying to improve their IELTS writing scores. There were three sub-groups: 85 participants attending IELTS preparation courses, 331 in the pre-sessional EAP course, and 60 in combination courses. All participants were asked to take the IELTS grammar/vocabulary tests at the beginning and end of their 4-to-14-week courses. Questionnaires examining participant and process variables such as learner background, motivation, class activities, and learning strategy use were completed after the pre and post tests. Inferential statistics were adopted and revealed “no clear advantage for focused test preparation” (p. 75). In addition, score gains were found primarily between two groups of learners: those who planned to take the test again, and those who had low initial writing test scores. “Washback to the learner (possibly in the form of motivation to succeed) rather than washback programme” (p. 93) has more to do with the improvement in students’ test scores. These findings have two implications: first, as indicated by Green (2007), test-driven instruction does not necessarily raise students’ scores. A more beneficial way to improve students’ scores may be to integrate material covered on the test with regular teaching. Second, concerning this point, intentions for taking the test need to be clear to both students and teachers to foster English learning.

Shih (2007) – GEPT as an exit requirement at two technical colleges in Taiwan

This study compares one private technical college in Taiwan which requires English majors to pass the elementary level of the General English Proficiency Test (GEPT) with a similar private technical college which has no such graduation requirement. The GEPT was commissioned by Taiwan's Ministry of Education in 1999 and is a criterion-referenced test that reputedly measures writing, speaking and listening skills.

Interviews with 2 department heads, 6 teachers, 30 students, and 3 family members were conducted. Observations were made for a semester in test-preparation classes or in classes that taught skills tested on the GEPT. Departments' policies regarding the GEPT exit requirements were also reviewed. Findings indicated that the GEPT had elicited a varying but minor impact on learners at both schools, although a slightly higher degree of washback was found at the school with exit requirements. In addition, Shih generated a new washback model of students' learning, as illustrated in Figure 1. This model includes extrinsic, intrinsic, and test factors to help depict the complexity of learning washback.

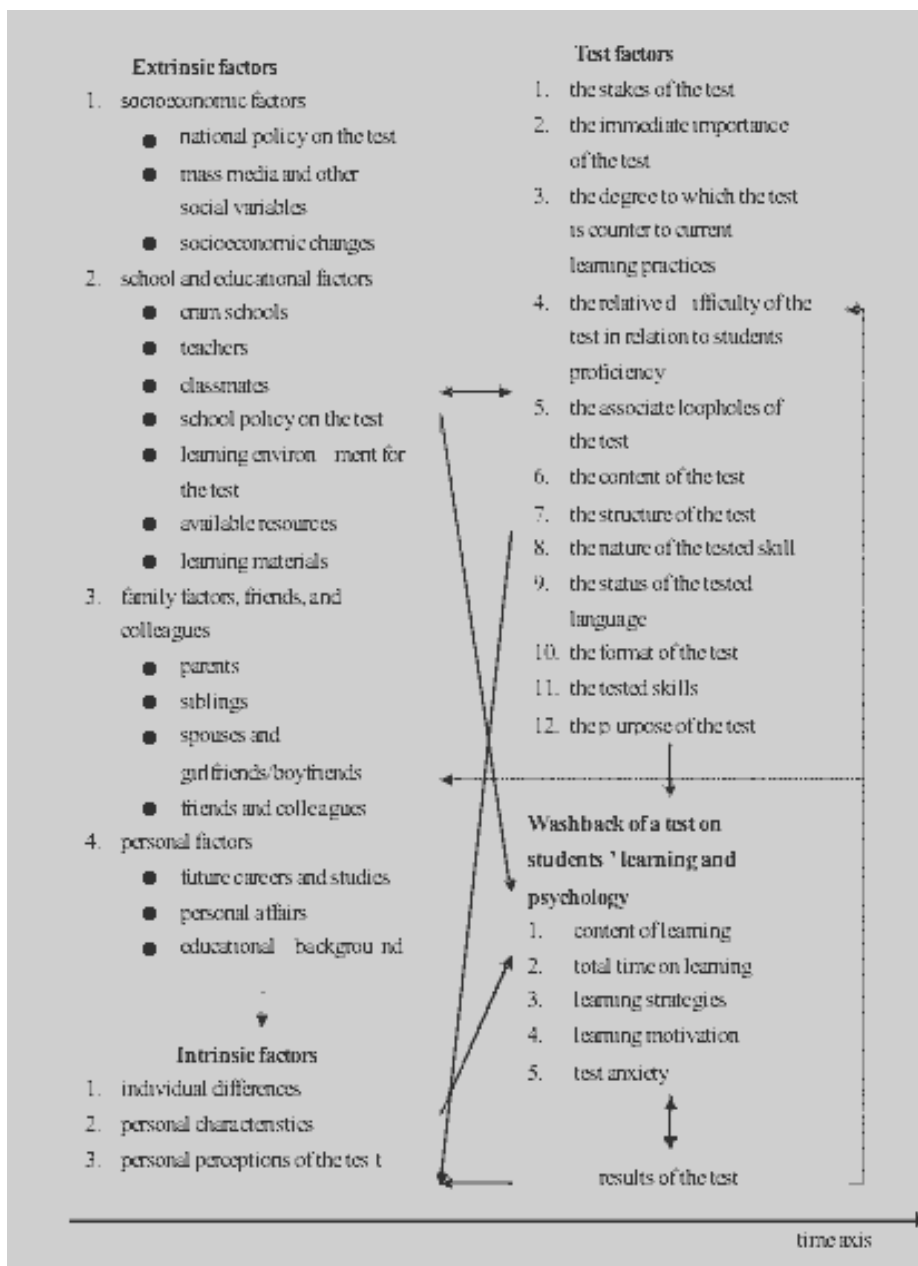


Figure 1. A proposed washback model of students' learning (Shih, 2007, p. 151).

Critical Discussion

Whereas these five studies all made significant contributions to different aspects of washback, perhaps their methods and results should be reconsidered carefully. At that point we will be in a better position to suggest avenues for further research.

Shohamy et al.'s study (1996)

Shohamy's study investigates test washback in a holistic way by looking at the self-reported influences of tests within classroom settings as well as on policy-makers, which has contributed a great deal to understanding washback from a macro point of view. As McNamara & Roever (2006), Messick, (1989) and Pennycook (1990) suggest, the influences of tests do not take place solely in the classroom; their ramifications are also social and political. Bachman & Palmer (1996) also claim that when exploring the broad phenomenon of washback, both micro effects in the classroom and macro effects on educational systems and society at large have to be examined. In light of this, Shohamy et al. shed useful information about test washback on policy-makers.

However, Shohamy et al.'s study has two significant limitations.

First, the fact that actual classroom observations were not included should lead us to question their findings, since what teachers claim they would do in class may vary from what they actually do. As stated by Cheng (2005) and Wall (2005), if we wish to know whether an exam can bring about changes in classroom teaching and learning, we must first examine the classroom itself, since that is where most teacher/student interaction occurs.

Second, their sample size (25 Israeli high school teachers and 112 students) was likely too small to warrant generalization. Simply stated, such a small size lacks statistical power. This may undermine the applicability of the research to larger contexts.

Alderson & Hamp-Lyons's study (1996)

Two points in Alderson and Hamp-Lyons's study are particularly strong. First, they incorporated an observational component in their study rather than relying solely on self-reports. Second, they used laughter as one barometer of the classroom atmosphere. Other studies by Cheng (2001), Hayes and Read (2004) and Watanabe (1997) have also considered laughter as a classroom variable.

One limitation of the study by Alderson and Hamp-Lyons is that they did not include questionnaires. Wall (2005) suggests that it is usually more difficult to reach a large number of respondents solely through observations and interviews; it is therefore useful to also supplement observations and interviews with questionnaires to explore the nature of washback effects (Bailey, 1999; Cheng, 2005; Wall, 2005).

A second limitation of Alderson and Hamp-Lyons's study was their choice of participants. Alderson and Hamp-Lyons point out that the TOEFL® affects both what and how teachers teach, but the effect differs considerably from teacher to teacher. However, given the varying backgrounds and amounts of teaching experience of the participants (a material developer with seventeen years of experience versus a first-time material teacher who has taught a TOEFL® class only once), the disparities of the effects come as no surprise (Saif, 1999). It would be worthwhile to determine whether those effects are similar among teachers with comparable backgrounds.

A third concern about the study by Alderson and Hamp-Lyons is that they dealt with washback primarily from teachers' perspectives, barely addressing students' points of view. To better understand how washback occurs within the classroom, researchers also need to investigate changes in students' motivations, learning styles, and learning strategies. Wall (2000) contends that many washback studies do not investigate learning outcomes, so it is necessary to address whether washback from exams affects learning, and if so, how. After all, preparation courses invariably claim to improve students' scores, but do they actually succeed?

One final concern about Alderson and Hamp-Lyons's study is that they did not make it clear what – if any – student score gains occurred. Although some studies (Hayes & Read, 2004) have shown that preparatory or intensive classes may not significantly affect score gains, it may be worthwhile to compare pre- and post-test scores between TOEFL® and non-TOEFL® classes. Moreover, a range of factors are found to be linked to score improvement, such as student personality, motivation, and exposure (Elder & O'Loughlin, 2003).

Cheng's study (1999, 2004)

As Cheng (2001), Alderson and Wall (1993), and Watanabe (1997) suggest, washback is a complex phenomenon that involves a variety of intervening variables such as tests, test-related teaching, learning and the perspectives of stakeholders. Given that complexity, washback studies often involve "naturalistic", "observational" and "descriptive" elements, so many washback studies utilize "survey research" (Hawkey, 2006, p. 32) approaches, especially in their use of questionnaires, interviews, and observation. Cheng uses all aforementioned methods and also adopted inferential statistics to analyze the differences in teacher perspectives of the old and new HKCEEs, thereby making the findings more convincing. Cheng contributes to the few washback studies by using both quantitative and qualitative methods.

Given the scarcity of baseline washback studies, Cheng's study is valuable because it attempts to gauge the effects of the new examination. Bailey (1996) claimed that the difficulty of conducting washback studies includes the problem of finding out what, if any, washback can be linked by evidence to the introduction and the use of tests.

Cheng's baseline study, by focusing on what occurred before the administration of the test and making a comparison between classroom activities and teachers' perspectives under the syllabus of both the old and new HKCEEs, helps us better understand what has changed. Qi (2004) points out that washback studies usually suffer from a lack of data collected before the test was first introduced. Cheng has thus provided us with a good starting point for more research.

However, a longitudinal study with a longer timeframe than the one used by Cheng might shed better light on the effects of the new HKCEE. As Messick (1989) claims, the effects of tests on societies and educational systems only becomes apparent after a while.

Green's study (2007)

Ross (2005) has pointed out that in a number of existing studies, the methods of research focus on the opinions that participants hold regarding washback effects. No significant effort is made to evaluate the actual results, which are the focal point of policy makers, as opposed to "perceptions of success" (p. 462). In light of this, Green's study is worthwhile in bettering our understanding of how washback influences learning outcomes. Regarding test washback, Green states:

if it is more generally found to be the case that 'teaching to test' is no more effective in boosting test scores than teaching the targeted skills, this will have profound implications for the relationship between teaching and testing. (p. 94)

Washback on learning outcomes is a complicated issue. It often is difficult to detect whether washback is due primarily to a test-prep course itself or other factors such as motivation, learning experience, and age. Green's research focuses on both participant and process variables, providing a comprehensive list for factors thought to influence learning outcomes. Those factors can be used in questionnaires for those interested in exploring similar topics.

However, it should be noted Green's study focused solely on learners' IELTS writing skill performance, and it would be more valuable if further investigation could be conducted to explore learning outcomes on listening, reading and speaking skills as well.

Shih's study (2007)

There has been comparatively little investigation of washback effects on students' learning processes. Watanabe (2004) states, "relatively well explored is the area of washback to the program, while less emphasis has been given to learners" (p. 22). Shih's study has therefore made a significant contribution to the work in this area.

In addition, it has provided us with a comprehensive list of extrinsic, intrinsic, and test factors that assist in the explanation of the intricacy of learning washback, while the previous washback theories of Alderson and Wall (1993) Bailey (1996)'s, and Hughes (1993) seem too simplistic in this respect. Shih's model also contributes to the explanation of how tests influence students' learning, especially applied to East-Asian contexts — foreign language education in Korea, Japan and Taiwan is remarkably similar.

One concern in Shih's model is that some items categorized as test factors share similarities such as the content, and test structure, test skills, as well as yet another distinguishing facet that Shih terms "the nature of the tested skills" which are all thought to have some influence on test performance. A more detailed explanation of how these items impact students' learning should also be provided. For example, Shih stated that test content influenced students' learning but did not indicate in what way. It is unclear whether students at the school where the GEPT is a graduation requirement spent more time listening to audio versions of test-preparation materials or not.

Another example regarding test impact is that Shih stated most students did not prepare for speaking test items because they did not know how to prepare for them. However, he did not clearly reveal the reasons for that. Was it due to the fact that no such speaking classes were offered or did students simply not have the motivation to practice their listening skills?

In addition, an explanation of how other factors in Shih's model such as the social-economic status of the examinees or status of the test in question might influence students' learning requires greater clarity as well. Do upper class parents tend to provide more financial support to children (paying for extra lessons at cram schools or purchasing more expensive test-preparation materials) than their counterparts? Also, to what degree is the GEPT promoted in the media via commercials and advertisements to encourage students' learning? The interaction between media blitzing and test performance remains unexplored.

Conclusions & Suggestions

“. . . most washback studies cover test effects on classroom settings or the educational contexts, while little attention is devoted to society at large.”

Each study cited here explores different aspects of washback. Shohamy et al. investigated how the status and stakes of tests influence teaching, learning and policy-makers' decisions on test use. Both Alderson and Hamp-Lyons and Cheng appraised washback on teaching by utilizing diverse instruments. Green's study examined learning outcomes, while Shih's study focused primarily on learning itself. A review of these studies reveals that most washback studies cover test effects on classroom settings or the educational contexts, while little attention is devoted to society at large.

However, we should remember that Hughes (2003) defines washback as the effects of testing not only on learners and teachers in a given educational context, but also on society at large. In addition, Bachman and Palmer (1996) have stressed that both classroom micro effects and macro social and educational system effects need to be examined. In light of this, overviews of washback should address both micro and macro levels.

Generated from the previous analysis of washback studies and the major washback models and current leading theories such as Alderson and Wall's fifteen washback hypotheses, Bailey's basic model of washback, and Hughes' trichotomy of washback, a proposed micro-and-macro washback model is presented in Figure 2.

Before examining this model however, let us briefly contrast it from previous washback models.

This model incorporates ideas from Hughes (1993, as cited in Bailey, 1999) in describing a trichotomy of test effects in terms of "participants", "process", and "product". Tests could affect teachers, students, administrators, materials writers, and publishers in terms of their perceptions, activities they engaged in, as well as the amount and quality of learning outcomes. Alderson and Wall (1993) propose fifteen washback hypotheses (Appendix 1) and illustrate some of their effects, from the most basic to the more specific, that tests might have on teaching and learning. For example, "A test will influence teaching/learning" (p. 120) and "Tests will have washback effects for some learners and some teachers, but not for others" (p. 121). Bailey (1996) combined the fifteen hypotheses from Alderson and Wall (1993) within the trichotomy of the backwash model proposed by Hughes (1993), and created the "basic model of washback" which appears in Appendix 2. Bailey distinguishes between "washback to the learner" (what and how learners learn and the rate/sequence and degree/depth of learning) and "washback to the program" (what and how teachers teach and the rate/sequence and degree/depth of teaching) to illustrate the mechanism by which washback works in actual teaching and learning contexts.

A common characteristic of these washback models is that they tend to highlight what washback looks like and who is affected, but do little to address the factors that contribute to the phenomenon. In other words, "process" is less understood than "participants" and "products". Besides, the products in these three models/hypotheses refer mainly to teaching and learning washback, not to the aspects of washback that might impact society.

The proposed model in Figure 2 aims to strive to represent a holistic balance of both micro-and-macro levels. Washback at the micro level is postulated to consist of teaching, learning, teaching material and scoregain effects, while washback at the macro level is postulated to consist of innovation and social dimension features. The different aspects of both levels are viewed as "products", in Hughes's (1993) term. "Tests + Participants", the first item in Figure 2, represents participants' (applying Hughes's terms) interactions with and perceptions toward tests, while "process", the second of Hughes's terms and the second item

in Figure 2, refers to the investigation of data derived from “Tests + Participants” intended to explain those products. In sum, to understand how these products evolve, an investigation of how participants themselves react toward tests must be conducted.

For example, to better understand why teachers change what they teach but not necessarily their methodology (Cheng, 1999, 2004, 2005) following the introduction of a test, their beliefs, perceptions of the test, and their levels of participation in its implementation may help us understand the phenomena of washback.

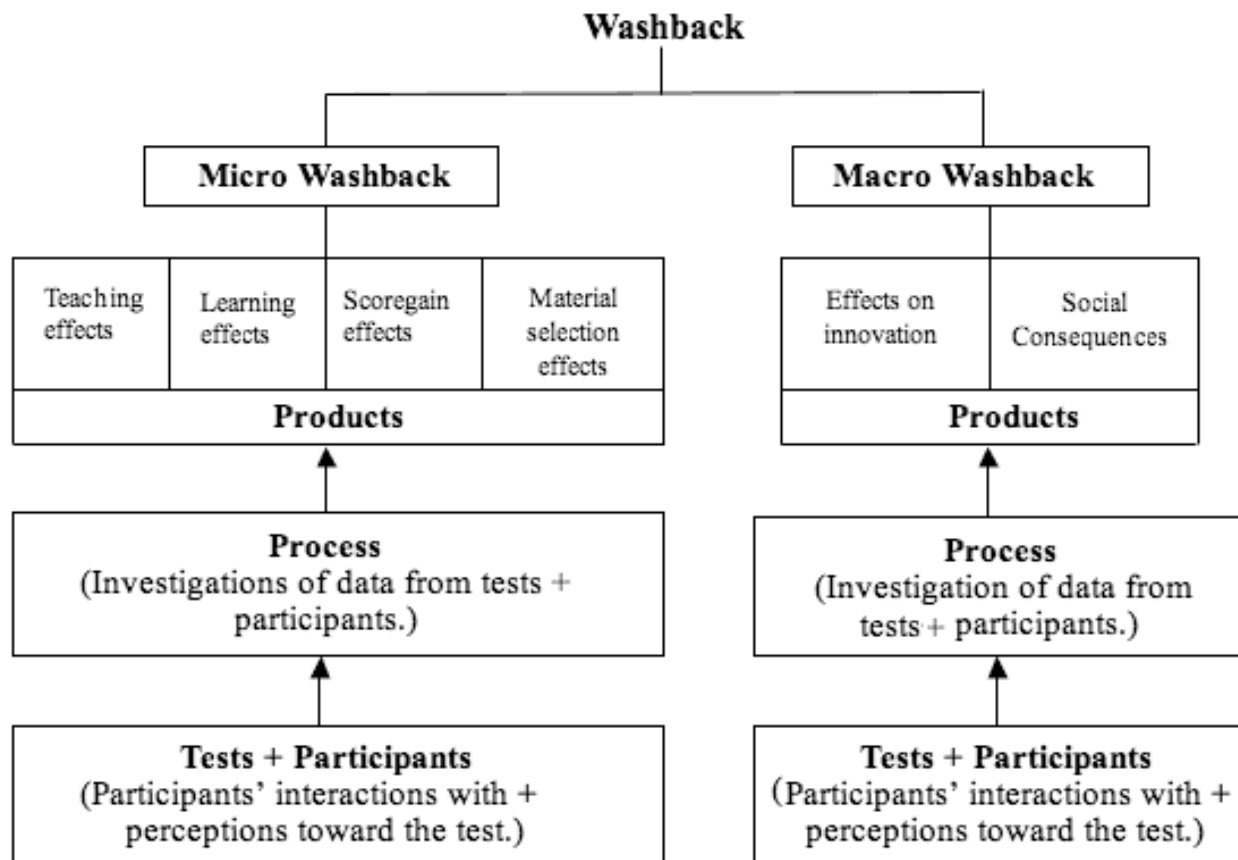


Figure 2. A proposed holistic model of washback

(Based on ideas of Hughes (1993), Bachman and Palmer (1996))

In summary, the model presented in Figure 2 investigates how three general phenomena interact on both the macro- and micro- levels. In addition, this model

“To gauge micro- and macro washback levels of washback, a triangulation of questionnaires, interviews, observations, pre-and-post tests, and document analysis need to be conducted.”

advocates a well-rounded investigation of washback that focuses not only on a given educational context but also society at large. To gauge micro- and macro washback levels of washback, a triangulation of questionnaires, interviews, observations, pre-and-post tests, and document analysis need to be conducted. This process involves many different

stakeholders such as teachers, students, administrators, policy-makers, family members and the general public.

References

- Alderson, C., & Hamp-Lyons, L. (1996). TOEFL Preparation Courses: A Study of Washback. *Language Testing* 13(3), 280-297.
- Bachman, F., & Palmer, S. (1996). *Language testing in Practice*. Oxford: Oxford University Press.
- Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2(1), 1-34.
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257-279.
- Bailey, K. M. (1999). *Washback in language testing. TOEFL Monograph Series*. New Jersey: Educational Testing Service.
- Cheng, L. (1999). Changing assessment: washback on teacher perceptions and actions. *Teaching and Teacher Education*, 15, 253-271.
- Cheng, L. (2001). Washback studies: methodological considerations. *Curriculum forum* 10(2), 17-32.
- Cheng, L. (2004). The washback effect of a public examination change on teachers' perceptions toward their classroom teaching. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 147-170). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Cheng, L. (2005). *Changing Language Teaching Through Language Testing: A Washback Study*. Cambridge: Cambridge University Press.
- Cronbach, L. J. (1988). Five perspectives on validation argument. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 3-17). Hillsdale, NJ: Erlbaum.
- Elder, C., & O'Loughlin, K. (2003). Investigation the relationship between intensive English language study and band score gain on IELTS. *International English Language Testing System Research Reports*, 4, 207-254.
- Green, A. (2007). Washback to learning outcomes: a comparative study of IELTS preparation and university pre-sessional language courses. *Assessment in Education*, 14 (1), 75-97.

Hawkey, R. (2006). *Impact theory and practice: Studies of the IELTS test and Progetto Lingue 2000*. Cambridge: Cambridge University Press.

Hayes, B., & Read, J. (2004). IELTS Test Preparation in New Zealand: Preparing students for the IELTS academic module. In L. W. Cheng, Y., and Curtis, A. (Ed.), *Washback in language testing: Research contexts and methods* (pp. 97-111). Mahwah, NJ: Lawrence Erlbaum Associates.

Hughes, A. (1993). *Backwash and TOEFL 2000*: University of Reading, England.

Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527-535.

McNamara, T., & Roever, C. (2006). *Language testing: the social dimension*. Oxford: Blackwell.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational Measurement* (pp. 13-103). New York: Macmillan.

Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3).

Pennycook, A. (1990). *The cultural politics of English as a world language*. London: Longman.

Qi, L. (2004). Has a high-stakes test produced the intended changes? . In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: research contexts and methods* (pp. 147-170). Mahwah, NJ: Lawrence Erlbaum Associates.

Ross, S. J. (2005). A book review of washback in language testing: research contexts and methods. In Cheng, L., Watanabe, and Curtis (eds.). *Applied Linguistics*, 26(3), 459-463.

Saif, S. (1999). *Theoretical and empirical considerations in Investigating Washback: A Study of ESL/EFL learners*. Unpublished Ph. D dissertation, University of Victoria, Canada.

Shih, C. (2007). A New Washback Model of Students' Learning. *The Canadian Modern Language Review*, 64 (1), 135-162.

Shohamy, E. (2001). *The Power of Tests*. Harlow: Pearson Education Limited.

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

Stoneman, B. W. H. (2006). *The impact of an exit English test on Hong Kong undergraduates: A study investigating the effects of test status on students' test preparation behaviours*. Unpublished Ph. D. Dissertation, The Hong Kong Polytechnic University, Hong Kong, China.

Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching*. Cambridge: Cambridge University Press.

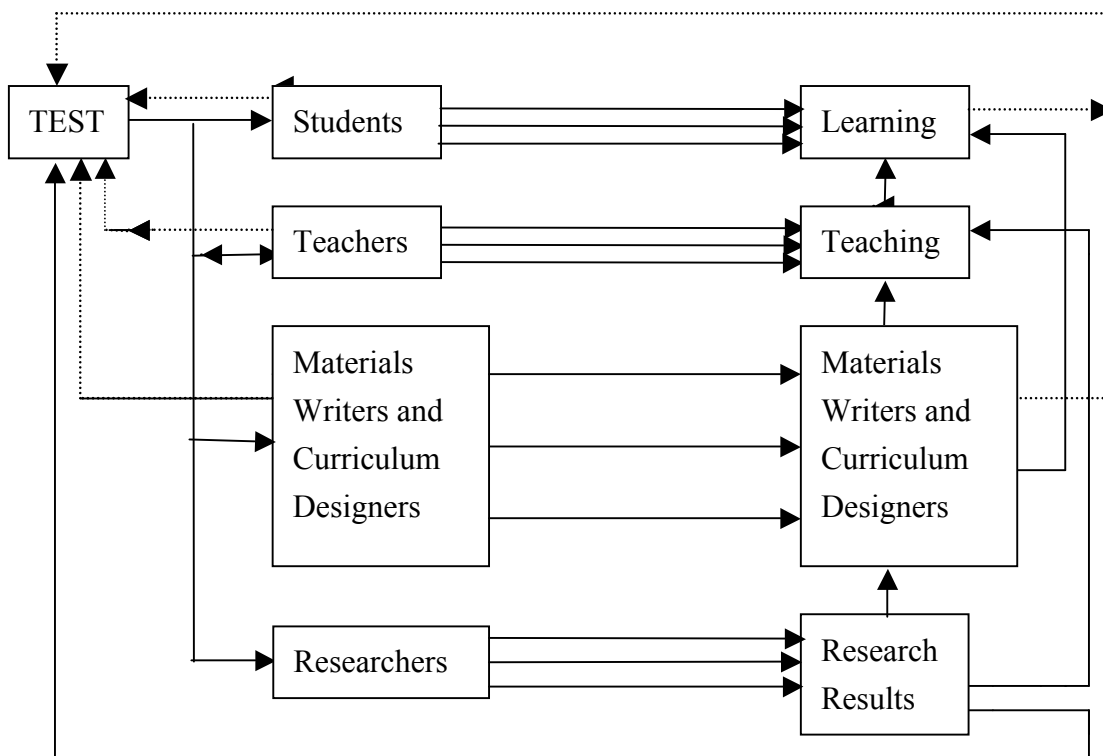
Wall, D., & Alderson, J. C. (1993). Examining washback: the Sri Lankan impact study. *Language Testing* 10(1), 41-69.

Watanabe, Y. (1997). Constructing a classroom observation scheme for the test impact research. *Sophia Linguistica*, 41, 297-313.

Appendix 1: Some Possible Washback Hypotheses Suggested by Alderson and Wall (1993, pp. 120-121)

1. A test will influence teaching.
2. A test will influence learning.
3. A test will influence what teachers teach; and
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; and
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 attitude towards the content, method, etc., of teaching and learning
12. Tests that have important consequences will have washback; conversely,
13. Tests that do not have important consequences will have no washback
14. Tests will have washback on all learners and teachers

Appendix 2: Bailey's Basic Model of Washback (1996, p. 264)



HTML: http://jalt.org/test/pan_1.htm / PDF: <http://jalt.org/test/PDF/Pan1.pdf>

Copyright © 2008 by PAN Yi-Ching & the Japan Association for Language Teaching