



THE AGA KHAN UNIVERSITY

eCommons@AKU

Institute for Educational Development, East Africa

AKU-East Africa

January 2007

Washback from language tests on teaching, learning and policy: Evidence from diverse settings

Pauline Rea-Dickins

Aga Khan University, pauline.rea-dickins@aku.edu

Catriona Scott

University of Bristol, United Kingdom

Follow this and additional works at: http://ecommons.aku.edu/eastafrica_ied

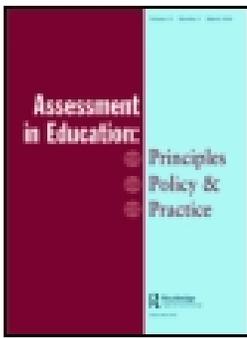


Part of the [English Language and Literature Commons](#), and the [Linguistics Commons](#)

Recommended Citation

Rea-Dickins, P., Scott, C. (2007). Washback from language tests on teaching, learning and policy: Evidence from diverse settings. *Assessment in Education: Principles, Policy & Practice*, 14(1), 1-7.

Available at: http://ecommons.aku.edu/eastafrica_ied/46



Washback from language tests on teaching, learning and policy: evidence from diverse settings

Pauline Rea-Dickins & Catriona Scott

To cite this article: Pauline Rea-Dickins & Catriona Scott (2007) Washback from language tests on teaching, learning and policy: evidence from diverse settings, *Assessment in Education: Principles, Policy & Practice*, 14:1, 1-7, DOI: [10.1080/09695940701272682](https://doi.org/10.1080/09695940701272682)

To link to this article: <http://dx.doi.org/10.1080/09695940701272682>



Published online: 25 Apr 2007.



[Submit your article to this journal](#)



Article views: 952



[View related articles](#)



Citing articles: 4 [View citing articles](#)

EDITORIAL

Washback from language tests on teaching, learning and policy: evidence from diverse settings

In his Editorial for Volume 10, Number 2 of *Assessment in Education*, Stobart (2003) draws our attention to the fact that all assessment has consequences, some of which are intended, others unintended. The authors in that issue of the journal addressed test impact from a general educational perspective, evidencing impact in terms of large-scale testing at national level, of how teacher evaluation may lead to a narrowing of the construct of teachers' professional knowledge and skills, as well as impact from the perspective of students.

Within applied linguistics, there is an equal concern with the consequences of assessment, and the effects that tests and examinations have on educational policy, processes and on a range of different stakeholders, i.e. those affected by tests both directly and indirectly. In the last decade, an increasing number of empirical studies have been undertaken by language testing and assessment researchers. There is, thus, a growing body of literature in both these fields. Whilst there is some evidence of language assessment researchers drawing on work in educational assessment (e.g., Wall & Horák, and Scott this volume), there appears to be little real communication between the education specialist and the applied linguist. One exception to this is the Cheng & Watanabe (2004) publication that provides a review of research in both language and general education, and brings together findings from a range of research settings, from general education in the US (Stecher *et al.*, 2004), to language education and the effects of International English Language Testing System (IELTS) (Cambridge Assessment, <http://www.cambridgeesol.org/index.htm>; <http://www.ielts.org/>) preparation courses in New Zealand (Hayes & Read, 2004) (see Book reviews, this issue). Given, too, that assessment has been advocated by policy-makers in diverse national contexts as an effective lever for change in raising attainment, both general education and language education researchers share interests in common regarding concerns of validity referenced to specific target audiences of test-takers, and tests as policy instruments, as well as to issues of test fairness and the ethical use of assessments. In other words, investigating the impact of our assessment practices is common to both research domains.

A considerable body of literature on impact in educational assessment centres on high-stakes testing in the US where state-mandated assessment has been and continues to be a key component of the drive to raise attainment since the publication of *A Nation At Risk* in 1983 (National Commission on Excellence in Education, April 1983) (e.g., Frederiksen, 1984, Airasian, 1987, Smith *et al.*, 1989; Haladyna *et al.*, 1991; Herman & Golan, 1993; Amrein & Berliner, 2002; Cimbricz, 2002; Braun, 2004). Interestingly, there is relatively little research into washback from national curriculum testing in England (cf. Broadfoot *et al.*, 2000) where educational reform aimed at raising attainment has involved the introduction of target setting and national testing at ages 7, 11 and 14 in the form of Key Stage Tests. In Wales, however, following a thorough review of statutory assessment (Daugherty, 2004a, b), secondary school league tables were dropped in 2001 and national tests for 7-year-olds ended in 2002 (see Scott, this volume).

The bulk of the empirical research into washback in the domain of language assessment also relates to high-stakes examinations in countries where English is used as a foreign or second language in, as examples, Sri Lanka (Wall & Alderson, 1993), Hong Kong (Andrews, 1995; Cheng, 1997, 1998, 1999; Andrews *et al.*, 2002) and Israel (Shohamy *et al.*, 1996). Much of the recent research has been framed within the context of higher education, focused specifically on two major and international language proficiency examinations, namely the International English Language Testing System (IELTS) and the Test of English as a Foreign Language (TOEFL™, <http://www.ets.org/portal/site/ets/menuitem.3a88fea28f42ada7c6ce5a10c3921509/?vgnnextoid=85b65784623f4010VgnVCM10000022f95190RCRD>)—see, for example, Green (2004), Hayes & Read (2004). Whilst the initial focus of washback studies centred on large-scale or international tests, the focus of the individual studies varied. In language education, Alderson & Wall (1993) may be seen to have led the agenda for washback research, with their fifteen ‘washback hypotheses’. These hypotheses and Bailey’s (1996) model of washback have been highly influential in providing frameworks for research into washback. Bailey’s model develops Hughes’ (1993) tripartite distinction between participants, process and products, and illustrates not only the way in which a test may affect products (e.g., learning, teaching, materials and research findings) through the participants and the processes they engage in, but also how these may in turn provide feedback resulting in changes to the tests. Further, a key facet of research into washback since Alderson & Wall’s (1993) seminal paper has been on classroom processes, in other words, what happens in the classroom (see, e.g., Wall & Alderson, 1993; Andrews, 1995; Alderson & Hamp-Lyons, 1996; Shohamy *et al.*, 1996; Cheng, 1997, 1999). In addition, there has been increasing interest in issues of equity, fairness, and the role of testing as a tool in the hands of policy-makers (see, e.g., Shohamy, 1997, 1998, 2001).

As Alderson (2004) himself notes in the foreword to Cheng & Watanabe (2004), research into washback has moved on since the early/mid-1990s. We now know that washback does exist as a phenomenon, but that it is extremely complex. It not only varies according to the test, but also in terms of contextual factors and participants’ responses. Burrows (2004) provides a useful set of models tracking the shift in views

of washback. The first represents the traditional view of washback, illustrated by a linear model with a single response by teachers to the introduction of a test. However, as Taylor (2006) points out, it is naïve to suppose that such a model works, i.e. that the introduction of a test or changes to a test can provoke desired changes in teaching and learning. It was Morrow (1986) who first spoke of ‘washback validity’ to describe the quality of the relationship between a test and teaching. Burrows’ (2004) second model shows individual teacher responses to a test, which reflects the view of washback held in the 1990s, but her research and other studies (e.g., Alderson & Hamp-Lyons, 1996) suggest another model, i.e. that there are patterns of teacher response. As Scott (2005) points out, this is a model which could be replicated by including the responses of other participants in the teaching process (e.g., the test candidates), adding to the complexity of the picture.

Reflecting awareness of the potential for effects that go beyond the classroom, a distinction is commonly made between ‘washback’, the effects on teacher and learner behaviour in the classroom, and ‘impact’, the wider effects of testing, e.g., on teaching materials, educational systems and wider society (Wall, 1997; Weir & Milanovic, 2003; Taylor, 2005; 2006). Indeed, washback can be considered to be one aspect of impact (Taylor, 2005), the former being micro-level and the latter macro-level effects (Weir & Milanovic, 2003). Whilst the term ‘washback’ (or ‘backwash’) is commonly used in language testing and applied linguistics, in educational research, it is more usual to refer to the same phenomenon as the ‘effects’ or ‘consequences’ of testing, incorporating notions of both ‘washback’ and ‘impact’.

Empirical research into washback, such as the examples cited above, has shown that there are features that are shared by both education and language testing. From a language assessment perspective, there is greater evidence of washback in terms of changes in the content of teaching than in methodology although, arguably, the latter may take longer to change and to be ‘seen’ (e.g., Cheng, 1997). The influence of tests on student learning is equally unpredictable, depending as it does on many factors, including teacher behaviours, but there is the potential for considerable impact in terms of affective factors. This last has been the focus of a recent study into the impact of the IELTS scores on student identity, learning and progression on a Master’s level programme at an English university (Rea-Dickins *et al.*, 2007). It took as its starting point that the institutional use of IELTS for university admissions reflects an implicit claim for a student’s language development and growth. The extent to which such potential is realized, or not, was considered a consequential validity issue relating to the IELTS examination, which was largely investigated through students’ ethnographic accounts of their experiences of postgraduate learning through the medium of English.

This special issue of *Assessment in Education* includes papers that focus on a variety of assessment mechanisms that have important consequences, from different perspectives in terms of their effects on learner (i.e. test-taker) performance and making judgements and predictions about students’ future attainment, on teaching and learning, in promoting pupil and parental knowledge and understandings about national testing systems, and, more widely, on how tests can exert a ‘reactionary grip’

on national policy. One paper also problematizes from a methodological perspective how washback may be researched.

The importance of investigating the consequences of large-scale school tests for different groups of learners is the focus of the contribution of Janna Fox and Liying Cheng. Their study relates to the Ontario Secondary School Literacy Test, validated on English first language (L1) populations, and investigates washback from this test on test-takers who were not first language speakers of English (L2). The accounts gathered from both L1 and L2 students reveal important differences in how they perceived the test's constructs and suggest a gap between what is valued as literacy on the test and what is valued in classroom literacy development practices, thus raising some concern about the test's consequential validity and its fairness for English L2 speakers, estimated to form over a quarter of the general school-age population in urban K–12 systems across Canada (Roessingh, 1999).

Catriona Scott also focuses on school-age learners and reports in her research on washback from statutory National Curriculum assessment on English L2 learners working in mainstream primary classes in England. Taking a stakeholder perspective to investigating washback, she explores the multiple purposes of the tests in relation to perceptions of the stakes of the tests and their impact on learners, teachers and the school. On the basis of interviews with teachers, parents and learners, she argues that, whilst the tests are not without value, particularly to the school in terms of tracking progress, their usefulness is limited in relation to providing meaningful information, particularly to parents. She therefore questions the balance between the merits of the tests and the stakes and potential for negative impact on schools, teachers and especially the learners.

Qi Luxia's research centres on a particular language skill, investigating the effects of a specific writing task in the National Matriculation English Test, a high-stakes test in the secondary school context in China. Her research involved a range of stakeholders: test constructors, teachers and students. Through interviews, classroom observation and questionnaires, she illustrates the potential for a discrepancy between the washback intended by test designers and actual washback in the classroom. She argues that the high-stakes nature of the test may undermine the implementation of intended washback on teaching as a result of teachers' and learners' desire to gain high test scores and their interpretation of how best this might be achieved. She concludes that high-stakes testing is not in itself an effective means of implementing change at the classroom level and that other measures are also required, namely teacher training and greater dialogue between test constructors and teachers.

The next two contributions to this special issue relate to major English language proficiency examinations: the TOEFL test (Educational Testing, Princeton, USA) and the IELTS (Cambridge Assessment, www.cambridgeassessment.org.uk). In preparing to take these major language proficiency examinations, a significant number of prospective test-takers enrol in English language programmes, some of which are dedicated to test preparation whilst others have a broader focus in the development of language skills required for effective university study. Anthony Green's study investigates the effects of three different programme types on the

development of students' writing skills and whether, in particular, dedicated test preparation classes give students an advantage in improving their scores on the writing sub-test of the IELTS.

The TOEFL-related research reported by Dianne Wall and Tania Horák has to do with selecting appropriate methodologies for researching washback. Drawing on a model designed to examine the effects of educational innovation, they use a range of approaches—documentary analysis, interviews, and classroom observation—in the development of a baseline study to establish the intended washback from a test on classroom pedagogy, and to explore characteristics of the users and the system in which they operate which might affect washback. They highlight the key features of their baseline study and the process of instrument development which could be applicable in other testing contexts. They point to the value of baseline studies in informing test designers of contextual factors relevant to the design and implementation of tests and in establishing a basis for exploring the link between intended impact and the eventual consequences of the test.

From the five studies above, we learn about investigations of washback from tests on: individual test-takers and the fairness of a high-stakes test for a particular group of students; on different stakeholders affected by national tests; on how a test may be influential in effecting pedagogical change; and on the relationship between test preparation practice and future test performance. The implications for policy, generally, and language education policy, in particular, are considerable. We therefore close this special issue with a paper by Elana Shohamy who problematizes how language tests are used as a mechanism in the manipulation of language education policies and policy control. Through her analysis and different examples, she identifies how language tests may present a stranglehold in terms of determining the prestige of language in an education system, the status and hierarchy of those languages, and also how they may function to suppress linguistic diversity through standardizing and perpetuating linguistic correctness and homogeneity. She reminds us of the immense power that examinations and tests may have in the hands of policy-makers. In these respects we can see the potential for language tests to represent a reactionary force within Applied Linguistics, a potential which needs to be resisted.

In preparing this Special Issue, our aim has been to showcase language focused research studies in the area of washback and to provide an opportunity for testing and assessment communities to draw on research into washback to inform their understanding of the impact of tests and provide a basis for discussion of the efficacy of assessment as a tool in improving teaching and learning. 'Washback', as represented in these papers, can be viewed as a context-specific shifting process, unstable, involving changing behaviours in ways which are difficult to predict. Rather than simply being an aspect of 'impact', washback perhaps follows from impact, equally unpredictable and changeable, but not necessarily malleable by external agency (e.g., test designers or policy-makers). Stobart (2003) maintains that given the inevitable consequences of testing and assessment 'The task is to make these as constructive as possible, particularly for those who are assessed' (p. 140). It is our view that this necessitates greater dialogue between language and education researchers and in this

volume we have brought together papers from a range of settings, and with different emphases, aimed at stimulating such a dialogue.

Pauline Rea-Dickins

Catriona Scott

Graduate School of Education, University of Bristol

References

- Airasian, P. W. (1987) State mandated testing and educational reform: context and consequences, *American Journal of Education*, 95, 393–412
- Alderson, J. C. (2004) Foreword, in: L. Cheng & Y. Watanabe, with A. Curtis (Eds) *Washback in language testing: research contexts and methods* (Mahwah, NJ and London, Lawrence Erlbaum Associates), ix–xii.
- Alderson, J. C. & Hamp-Lyons, L. (1996) TOEFL preparation courses: a study of washback, *Language Testing*, 13(3), 280–297.
- Alderson, J. C. & Wall, D. (1993) Does washback exist?, *Applied Linguistics*, 14(2), 115–129.
- Amrein, A. L. & Berliner, D. C. (2002) High-stakes testing, uncertainty, and student learning, *Education Policy Analysis Archives*, 10(18), 1–61. Available online at: <http://epaa.asu.edu/epaa/v10n18.html> (accessed 16 May 2002).
- Andrews, S. (1995) Washback or washout? The relationship between examination reform and curriculum innovation, in: D. Nunan, V. Berry & R. Berry (Eds) *Bringing about change in language education* (Hong Kong, University of Hong Kong), 67–82.
- Andrews, S., Fullilove, J. & Wong, Y. (2002) Targeting washback: a case study, *System*, 30, 207–223.
- Bailey, K. M. (1996) Working for washback: a review of the washback concept in language testing, *Language Testing* 13(3), 257–279.
- Braun, H. (2004) Reconsidering the impact of high-stakes testing, *Education Policy Analysis Archives*, 12(1). Available online at: <http://epaa.asu.edu/epaa/v12n1> (accessed on 25 June 2004).
- Broadfoot, P., Osborn, M., Planel, C. & Sharpe, K. (2000) *Promoting quality in education: does England have the answer?* (London, Continuum).
- Burrows, C. (2004) Washback in classroom-based assessment: a study of the washback effect in the Australian adult migrant English Program, in: L. Cheng & Y. Watanabe, with A. Curtis (Eds) *Washback in language testing: research contexts and methods* (Mahwah, NJ and London, Lawrence Erlbaum Associates), 113–128.
- Cheng, L. (1997) How does washback influence teaching? Implications for Hong Kong, *Language and Education*, 11(1), 38–54.
- Cheng, L. (1998) Impact of a public English examination change on students' perceptions and attitudes toward their English learning, *Studies in Educational Evaluation*, 24, 279–301.
- Cheng, L. (1999) Changing assessment: washback on teacher perceptions and actions, *Teaching and Teacher Education*, 15, 253–271.
- Cheng, L. & Watanabe, Y., with Curtis, A. (Eds) (2004) *Washback in language testing: research contexts and methods* (Mahwah, NJ and London, Lawrence Erlbaum Associates).
- Cimbricz, S. (2002) State-mandated testing and teachers' beliefs and practice, *Education Policy Analysis Archives*, 10(2). Available online at: <http://epaa.asu.edu/epaa/v10n2.html> (accessed on 25 June 2004).
- Daugherty, R. (2004a) *Learning pathways through statutory assessment: Key Stages 2 and 3: interim report of the Daugherty Assessment Review Group* (Cardiff, Welsh Assembly Government).
- Daugherty, R. (2004b) *Learning pathways through statutory assessment: Key Stages 2 and 3: final report of the Daugherty Assessment Review Group* (Cardiff, Welsh Assembly Government).

- Frederiksen, N. (1984) The real test bias: influences of testing on teaching and learning, *American Psychology*, 39, 193–202.
- Green, A. (2004) Can placement tests inform instructional decisions?, *Language Testing*, 21(4), 467–494.
- Haladyna, T. M., Nolen, S. B., & Haas, N. S. (1991) Raising standardized achievement test scores and the origins of test score pollution, *Educational Research*, 20(5), 2–7.
- Hayes, B. & Read, J. (2004) IELTS test preparation in New Zealand: preparing students for the IELTS academic module, in: L. Cheng & Y. Watanabe, with A. Curtis (Eds) *Washback in language testing: research contexts and methods* (Mahwah, NJ and London, Lawrence Erlbaum Associates), 97–111.
- Herman, J. L. & Golan, S. (1993) The effects of standardized testing on teaching and schools, *Educational Measurement: Issues and Practice*, 12(4), 20–25, 41–42.
- Hughes, A. (1993) *Backwash and TOEFL 2000*. Unpublished manuscript, University of Reading.
- Morrow, K. (1986) The evaluation of tests of communicative performance, in: M. Portal (Ed.) *Innovations in language testing* (London, NFER/Nelson), 1–13.
- National Commission on Excellence in Education (April 1983) *A nation at risk: the imperative for educational reform*. Available online at: <http://www.ed.gov/pubs/NatAtRisk/index.html> (accessed 9 January 2007).
- Rea-Dickins, P., Kiely, R. & Yu, G. (2007) *Student identity, learning and progression: with specific reference to the affective and academic impact of IELTS on 'successful' candidates. Vol. 7. IELTS Impact Studies* (Australia, IELTS Australia and British Council).
- Roessingh, H. (1999) Adjunct support for high school ESL, *TESL Canada Journal*, 17(1), 72–87.
- Scott, C. (2005) *Washback in the UK primary context with EAL learners: exploratory case studies*. Unpublished doctoral dissertation, University of Bristol.
- Shohamy, E. (1997) Testing methods, testing consequences: are they ethical? Are they fair?, *Language Testing*, 14(3), 340–349.
- Shohamy, E. (1998) Critical language testing and beyond, *Studies in Educational Evaluation*, 24(4), 331–345.
- Shohamy, E. (2001) *The power of tests: a critical perspective on the uses of language tests* (Harlow, Longman).
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996) Test impact revisited: washback effect over time, *Language Testing*, 13(3), 298–317.
- Smith, M. L, Edelsky, C., Draper, K., Rottenberg, C., & Cherland, M. (1989) *The role of testing in elementary schools* (Los Angeles, CA, Center for Research on Educational Standards and Student Tests, Graduate School of Education, UCLA).
- Stecher, B., Chun, T. & Barron S. (2004) The effects of assessment-driven reform on the teaching of writing in Washington State, in: L. Cheng & Y. Watanabe, with A. Curtis (Eds) *Washback in language testing: research contexts and methods* (Mahwah, NJ and London, Lawrence Erlbaum Associates), 53–69.
- Stobart, G. (2003) Editorial. The impact of assessment: intended and unintended consequences, *Assessment in Education*, 10(2), 139–140.
- Taylor, L. (2005) Washback and impact, *ELT Journal*, 59(2), 54–155.
- Taylor, L. (2006) The changing landscape of English: implications for language assessment, *ELT Journal*, 60(1), 51–60.
- Wall, D. (1997) Impact and washback in language testing, in: C. Clapham & D. Corson (Eds) *Encyclopedia of language and education: Vol. 7. Language testing and assessment* (Dordrecht, Kluwer Academic), 291–302.
- Wall, D. & Alderson J. C. (1993) Examining washback: the Sri Lankan impact study, *Language Testing*, 10(4), 41–69.
- Weir, C. J. & Milanovic, M. (2003) *Continuity and innovation: revising the Cambridge Proficiency in English Examination 1913–2002* (Cambridge, UCLES/Cambridge University Press).