

# TESTS AS DRIVERS OF CHANGE IN EDUCATION: CONTEXTUALISING WASHBACK, AND THE POSSIBILITY OF WASH-FORWARD

Gordon Allan\*

*British Council, 1-2 Kaguruzaka, Shinjuku-ku, Tokyo, 162-0825, Japan*

Received 26 February 2020

Revised 15 March 2020; Accepted 25 July 2020

**Abstract:** Whilst the primary aim of tests is to measure ability, it is not uncommon for tests to be deployed in education systems with the intention, at least in part, of driving change in educational practice by making demands that teachers and learners are expected to meet. Washback is one way by which teaching and learning practices may adapt to a new test, but it is not the only possibility and often fails to occur as intended. This paper seeks to draw together ideas from different sources to place washback in the context of other possibilities. The concepts of adaptive implementation and programmed implementation are taken from Henrichsen's hybrid model of the diffusion/implementation of innovation in education systems. Washback is shown to act in parallel to but distinct from programmed implementation. The picture is completed with van Lier's concept of wash-forward, first outlined in 1989 but subsequently neglected in the literature. Wash-forward is illustrated with an example from the implementation of the National Matriculation English Test (NMET) in China. The intention is to provide an easily visualised, refreshed and more complete perspective on the processes operating when a new test is introduced as part of a strategy aimed at driving changes in teaching and learning practices; a scenario which is very relevant to the current movement towards four-skills English testing in East Asia and around the world.

*Keywords:* Washback, wash-forward, programmed implementation, innovation in education, exam reform

## 1. Introduction: Tests as drivers of change in education

High-stakes tests are often introduced in the hope of driving educational change by eliciting positive washback (for examples see Wall & Alderson, 1993; Cheng, 2002; Qi, 2005; MEXT, 2014). In such cases, it is hoped that the demands of a new test will help to drive changes in teaching and learning practices to produce more effective learning outcomes. The literature, however, shows that the desired washback often fails to occur (e.g. Wall & Alderson, 1993; Qi, 2005). It is

therefore worth considering the mechanism by which tests might drive educational change, and what might be happening when the desired washback does not transpire.

The present paper first briefly reviews the concept of washback, then attempts to contextualise it by relating it to Henrichsen's hybrid model of the diffusion/implementation of innovation in education systems (Henrichsen, 1989). Finally, the concept of "wash-forward" (van-Lier, 1989, p.494) is added to make up a picture of two pairs of parallel processes acting from opposite sides to bridge the gap between the demands of a new test on one side and existing teaching and learning practices on the other.

\* Tel.: +81-90-9954-3184

Email: Gordon.Allan@britishcouncil.or.jp

**2. Washback**

At its simplest, washback (synonymous with backwash) has been defined as “the effect that tests have on learning and teaching” (Hughes, 2003, p.53). Buck (1988) views washback as “a natural tendency for both teachers and students to tailor their classroom activities to the demands of the test” (p.17), particularly high-stakes tests, noting that it “can be either beneficial or harmful”. Messick (1996) expands on this somewhat, defining washback as “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” (p.241). This “otherwise” is of course hypothetical, and usually gauged by a baseline study of what teachers and learners did prior to a test’s introduction. Since this baseline may vary between contexts and/or individuals, the same test may elicit different washback from different contexts and individuals (examples of washback varying between individuals are summarised in table 1). Washback is therefore not a property of a test *per se*, but a consequence of test use.

Table 1: Examples of washback varying between individuals in the same context

Study	Alderson & Hamp-Lyons (1996)	Watanabe (1996)
Comparison	Language proficiency classes vs. TOEFL preparation classes	Preparation classes for the English sections of two types of examinations: Entrance examination for a national university [grammar-translation (GT) oriented] Entrance examination for a private university [non-GT oriented]
Context	Specialised language institute for foreign students in the USA “whose students are regularly admitted to USA universities” (p. 283)	<i>Yobiko</i> (privately-run extra-curricular entrance exam preparation school) for Japanese students in central Tokyo
Differences in washback reported	Study observed two teachers in TOEFL prep and regular lessons. Both teachers showed some consistent differences between their TOEFL-prep and regular classes, e.g. greater teacher talk time, fewer opportunities for student talk, and less use of pairwork. However, the two teachers’ TOEFL-prep lessons also varied considerably in approach, e.g. use of metalanguage, test-practice in class time, and opportunities for discussion. In contrast to the two observed teachers, another teacher (new to TOEFL teaching, not observed) reported thinking that interaction was vital in a TOEFL prep class and described a number of interactive techniques she used in her TOEFL prep lessons.	Two teachers observed teaching exam preparation classes for both types of exam. One focused on translation and explaining structures for both exams, regardless of whether the exam was GT-oriented. The other varied his approach, translating and explaining more in the GT-oriented exam preparation lessons than in the non-GT-oriented exam preparation lessons.

More recently, Cheng (2005) defines washback as “an intended or unintended (accidental) direction and function of curriculum change on aspects of teaching and learning by means of a change of public examinations”, which is consistent with previous definitions in that washback refers to examinations changing teaching and learning practices, and not the reverse. Since washback has the potential to bring about change in educational practice, it is worth considering how it fits with a model of innovation in education.

### 3. Henrichsen’s hybrid model of diffusion of innovation in education

Henrichsen’s (1989) hybrid model of the diffusion/implementation of innovation in education systems follows the course of an innovation, from its antecedent state, through the process of its implementation, to the decision to adopt or reject and subsequent consequences. It remains a rich and comprehensive account of the factors influencing the success or otherwise of educational innovations.

Although washback has a role in the diffusion of innovation, it is not immediately obvious where it fits into Henrichsen’s (1989) model. It is a consequence of test use but does not appear in the consequences phase of the model because the model is concerned with the fate of the test (innovation) itself rather than

the test’s effect on other practices. Instead, interaction between the test and its user system is considered as part of the process phase, in the section on “inter-elemental factors”.

Henrichsen (1989) argues convincingly that mismatch between an innovation and its user system potentially threatens the successful adoption of the innovation. Where tests are deployed to drive change in educational practices, however, the motive force for change can only be generated by deliberately creating precisely such a mismatch, since a test that is well-aligned with existing practice provides no motivation for change. When a test makes demands that are not being met by the educational status quo, a gap opens between the test on one side and teaching and learning practices on the other. To bring the two back into harmony, this gap must be closed.

A full account of the nature of potential dissonance between the demands of a new test and existing educational practice is beyond the scope of the present paper, which seeks to deal with the topic at a conceptual level. As illustrated in the examples summarised in table 2, however, in general terms it is not uncommon for new tests to emphasise the practical use of English for communicative purposes where this is perceived to be deficient in existing English teaching and learning practices.

Table 2: Examples of dissonance between the demands of a new test and existing educational practices, with the intention of inducing change in education systems

Study	Wall & Alderson (1993)	Qi (2005)
Context	Introduction of a new O-level English exam, custom-made to reinforce new textbooks, in Sri Lanka, 1988.	Introduction of the National Matriculation English Test (NMET) in China, 1985.

Nature of dissonance between the demands of the test and existing teaching and learning practice	Prior to this, teaching tended to be structurally based and focused on general reading skills, with little opportunity for communication. The new O-level and accompanying textbooks introduced greater emphasis on reading and writing for a communicative purpose, a previously neglected aspect of both skills.	Prior to the NMET, English teaching was characterised by a focus on grammatical structures and rote-learning of vocabulary, with little attention to communicative skills. The NMET shifted the focus from formal linguistic knowledge to the practical use of language.
Intended change	The development of more practical English communication skills.	“...to replace the time-honored traditional method in China’s ELT with the widely accepted communicative approach.” (p.145)

Henrichsen (1989, p.92) cites Roberts-Gray & Gray (1983) to describe two processes for the resolution of dissonance between an innovation and its user system: adaptive and programmed implementation. Adaptive implementation involves altering the innovation to fit the users, and may include such processes as localising test content, for example to ensure that pictures

used in the test are recognisable to test-takers. Programmed implementation works in the opposite direction, seeking to change the user system to accommodate the innovation, an example of which might be a teacher training programme aimed at helping teachers to prepare students for a new test. Table 3 shows illustrative examples of programmed and adaptive implementation in practice.

Table 3: Examples of programmed and adaptive implementation from the Sri Lanka study (Wall & Alderson, 1993)

Study	Wall & Alderson (1993)
Context	Introduction of a new O-level English exam in Sri Lanka, 1988.
Adaptive implementation (aimed at changing the test to accommodate the test user system)	<ul style="list-style-type: none"> <li>The new exam was custom made to align with new textbooks. Test developers analysed the textbooks and drew up test specifications in consultation with the textbook writers. The new textbooks and new exam were therefore well-aligned and mutually supportive of each other.</li> <li>There was a conscious attempt to choose texts, topics, and authentic tasks that were relevant to Sri Lankan school children and their likely purposes for using English.</li> <li>Plans to employ continuous assessment to test oral skills were dropped because they proved practically and politically impossible.</li> </ul>
Programmed implementation (aimed at changing the test user system to accommodate the new test)	<ul style="list-style-type: none"> <li>The first textbook in the new series was accompanied by a Teacher’s Guide including guidance on the essentials of the approach teachers were expected to follow and how to use the new textbook material in the classroom.</li> <li>Teacher training efforts also accompanied the introduction of the new textbooks.</li> </ul>

Described thus, both adaptive and programmed implementation are active interventions to support the success of an innovation. Washback is distinct in that it is not an active intervention, but a consequence of a test's interaction with its educational context. Although washback acts in the same direction as programmed implementation (i.e. the users adapt to the innovation), it arises from teachers and learners as a response to the test. This view is consistent with that of Messick (1996), who is careful to maintain a distinction between washback and other drivers of change, such as training programmes or new textbooks. In addition, washback may be unpredictable and inconsistent (e.g. Watanabe, 1996) in contrast to what is implied by the term "programmed".

The relationship between programmed implementation and washback is an important one. As Cheng (2002) notes, while tests may provide the motivation for change, they do not provide the knowledge or skills required to enact that change. As teachers and learners respond to the demands of the test, any new knowledge or skills that may be required of them must be drawn from somewhere. A test can only elicit what teachers and learners have the capacity to provide. This may help to explain findings such as Wall and Alderson's (1993) observation that washback was evident in content taught but not teaching methodology.

The role of programmed implementation in creating the potential for positive washback is therefore crucial, for example via training to foster the development of new skills, or the development and provision of appropriate resources such as textbooks. As Cheng (2002) puts it:

The change to a new exam has informed teachers about what they might do differently, but it has not shown them how to do it.

The washback effect can be fully realized only when all levels of organizations in the educational system are involved. In this sense, there must be a genuine involvement of educators and textbook writers. A change in the final examination alone will not achieve the intended goal (Cheng, 2002, p.108).

In summary, it seems helpful to consider washback as parallel to but distinct from programmed implementation. This raises the question of whether there is any equivalent parallel to adaptive implementation. The literature suggests that there is.

#### **4. "Wash-forward" (van Lier, 1989)**

##### ***4.1 What is wash-forward?***

In his 1989 paper on the oral proficiency interview (OPI), Leo van Lier observes that:

By pushing for innovative testing techniques, particularly the OPI, proficiency advocates hope that a desirable washback effect will be created; in other words, if teachers and learners know that tests will demand both communicative ability and accuracy, the methodological focus of classroom work will change accordingly. This is something of an act of faith, of course, since it is also possible that classroom practices will prove so recalcitrant that they will force the OPI to shift in the direction of standard curricula. (van Lier, 1989, p.491)

The paper goes on to introduce the concept of "a possible 'wash-forward' (as opposed to washback) effect of methodological and curricular concerns carrying over into the rating" (van Lier, 1989, p.494). In other words, existing perceptions and/or practices may influence how a test construct is operationalised.

The example of possible wash-forward given by van Lier (1989) is that, in practice,

aside from their training and the descriptors given in rating scales, examiners may tend to over-rely on linguistic features that are given prominence in syllabuses (e.g. conditional forms), particularly when making decisions at the boundaries between bands/grades. Relating this to Henrichsen's (1989) hybrid model, wash-forward acts in the same direction as adaptive implementation by "modifying or adapting the innovation so that it is more easily assimilated into user practices and values" (Roberts-Gray & Gray, 1983, p.216).

It seems obvious that van Lier's (1989) concept of wash-forward must be a possibility, but apart from a brief mention by Johnstone (1994), it does not appear to have been followed up in the literature. It is not immediately obvious why not. When this paper was first presented at the AALA conference in Hanoi, 2019, it was suggested that perhaps wash-forward had been considered as a type of washback. This may be the case, since both washback and wash-forward are consequences of test use, but it goes against the definition of washback as "the effect that tests have on learning and teaching" (Hughes, 2003, p.53), consistent with other definitions (e.g. Buck, 1988; Messick, 1996; Cheng, 2005). Thus defined, washback cannot include its opposite, i.e. the effects that teaching and learning have on tests, so a distinct term is required, hence wash-forward.

"Wash-forward" (van Lier, 1989) is perhaps not a particularly intuitive term for the phenomenon it refers to. It was presumably chosen because it sounds like the opposite of washback, but a more clearly descriptive term may well be preferable, and it is even possible that the lack of such has played a role in the subsequent neglect of the concept. To avoid confusion, however, no alternative terms are

suggested here, and wash-forward is used throughout.

It is also worth noting that the suggested mechanism for wash-forward in van Lier's (1989) example is via the rating process. This mechanism can only influence assessment of the productive skills (speaking and listening), since there is no equivalent rating process for the receptive skills (listening and reading). Further research would be required to determine whether there is any other mechanism by which wash-forward might occur.

Despite the term and concept being neglected, however, if wash-forward is a real possibility we might expect examples of it to be found in the literature. Qi's (2005) account of the National Matriculation English Test (NMET) in China appears to document just such a possible occurrence in practice.

#### ***4.2 A possible example of wash-forward in the implementation of the NMET, China (Qi, 2005)***

Introduced in 1985, one of the aims of the English component of the NMET was to promote a shift away from traditional methods of English education towards a more communicative approach (Qi, 2005). As part of this effort, a communicative context (audience, purpose) was provided in the rubric for the writing task, and the marking criteria included appropriacy (Qi, 2005). However, the inclusion of appropriacy only lasted for the first six years, until 1990, before it disappeared from the marking criteria, subsequently reappearing only inconsistently and/or indirectly.

In her study, Qi found that teachers and students did not see the specified communicative context as a basis upon which to choose an appropriate writing style. Instead,

they “interpreted the trait measured ... as the ability to extract the prescribed content from the input and put it down neatly in grammatically correct English sentences” (Qi, 2005, p. 157). In addition, NMET markers’ advice about improving scores, tended to emphasise handwriting, accuracy and the extraction of key points from input, but did not mention appropriacy (Qi, 2005).

Qi (2005) concluded that “appropriacy does not significantly affect the scores for writing in the NMET” (p.158), an apparent example of construct under-representation (Messick, 1996), not by design but due to the way the construct was operationalised. Qi (2005) attributed this in part to teachers lacking sufficient experience of English communication to judge appropriacy. She cites an example of a teacher recommending over-polite language as appropriate for a letter to a railway station lost property office: “Anyway, you should sound polite... It’s better if you write ‘I would be very much obliged if you could kindly help me to find it.’” (Qi, 2005, p. 159). This is consistent with concerns raised by Cheng (2002) that for change to be

enacted as desired, those enacting it require the knowledge and skills to do so.

So, despite the inclusion of appropriacy in the early marking criteria, it seems questionable whether that aspect of the construct was ever operationalised. In the absence of a sufficient base of the required knowledge and skills, it appears that educational practices may have proved “so recalcitrant” (van Lier, 1989, p.491) as to cause an effective alteration of the construct of the test; wash-forward as opposed to the intended washback.

Qi’s (2005) account reminds us that a high-stakes test is not an immovable object but is subject to the influence of the context(s) within which it is deployed.

## 5. Summary

The above has sought to place washback in the broader context of the interactions that may occur when a test is deployed with the aim of driving change in teaching and learning. A schematic representation of the four processes described is given in figure 1:

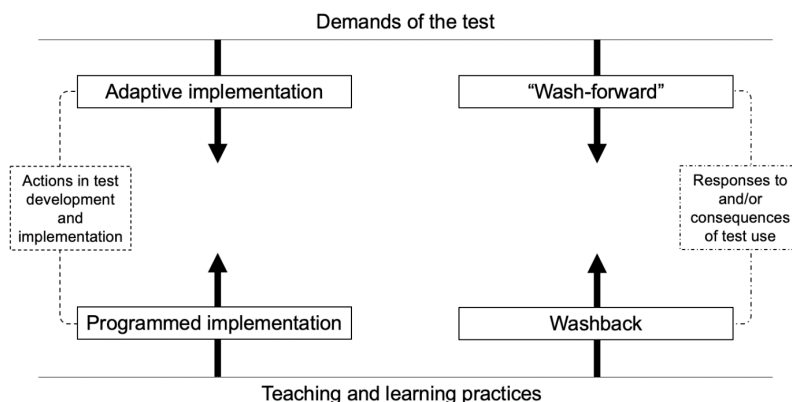


Figure 1: The processes by which dissonance between the demands of a test and the practices of teachers and learners might be resolved

The processes of programmed and adaptive implementation cited by Henrichsen (1989) are complemented by the well-established

concept of washback on the one hand, and van Lier’s (1989) concept of wash-forward on the other. Programmed and adaptive

implementation are considered as actions (interventions) taken by those implementing a new test, while washback and wash-forward are responses to and/or consequences of test use. It should be noted that, as with other consequences of innovation, washback and wash-forward may be immediate and/or delayed, direct and/or indirect, manifest and/or latent, functional and/or dysfunctional (Henrichsen, 1989, p.95).

## **6. Implications**

O'Sullivan (in press) argues convincingly that for an education system to work, assessment must be in harmony with the curriculum and all aspects of delivery. This is consistent with Henrichsen's argument that mismatch between an innovation and its user system is a threat to the success of the innovation. O'Sullivan's argument takes this further; in a comprehensive learning system, any mismatch between the key elements of assessment, curriculum, and delivery does not only threaten the mismatched element, but puts the whole system at risk (O'Sullivan, in press).

The present paper has sought to examine in detail one possible source of mismatch within a learning system, and place the widely studied phenomenon of washback in the context of other processes that may operate when such a mismatch occurs. Since the different parts of an effective learning system are interconnected, there are implications for a number of different stakeholder groups. Central to these is the importance of communication between various stakeholder groups, in terms that everyone can understand.

### **6.1 Implications for test developers**

Test developers need to work with curriculum planners and classroom practitioners to achieve a harmonious

system. They need to help inform teachers and learners about tests, but also need to research and understand the context in which the test will be used. Such understanding can inform adaptive implementation, bringing the test closer to the needs and potential of teachers and learners. It can also be helpful in identifying opportunities for programmed implementation to equip teachers with the knowledge and skills to help learners meet the demands of any new test, thereby helping to generate potential for positive washback.

### **6.2 Implications for education policy makers**

Qi's (2005) account of the NMET makes clear that if a test is not well-aligned with the knowledge and abilities of those implementing it, wash-forward is a possible consequence, which may undermine the test construct and its potential to drive desired learning outcomes. Given the expense involved in developing and implementing a high-stakes test, this might represent a very costly failure. Policy makers therefore need to work with test developers, teachers, and researchers to understand the demands of any new test, the capacity of teachers and learners to meet these demands, and the support required to make the system work.

### **6.3 Implications for teachers and learners**

Teachers and learners also share some responsibility for understanding assessment and how it relates to classroom practice. It is important that they understand what is being tested so that they can approach language learning and test preparation constructively. Developing the knowledge and skills required may involve engaging with training, asking questions, and being open to new ways of doing things. Equally, it may involve working with researchers to help them understand how a test is perceived and identify what support



may be needed to make it work in context. Just as it is important for test developers and policy makers to listen, it is equally important for teachers and learners to have a voice in the conversation, and to use it to feed back their perceptions and experience.

#### **6.4 Implications for researchers**

Researchers have an important role to play in facilitating effective communication between the elements of a learning system, by collecting and analysing data including different stakeholder perspectives. It is therefore important that researchers understand the interactions between the elements of the system.

The contextualisation in the present paper is a reminder that washback is not synonymous with the broader category of impact, but one aspect of impact. The relatively high profile of washback in language testing research has perhaps put it at risk of becoming a catch-all term occasionally (mis)used to refer to consequences of test use that are not covered by the clear and consistent definitions in the literature. For example, a recent review of “ten empirical studies of washback in language teaching conducted around the world between 2011 and 2018” (Ahmmed & Rahman, 2019, p.153), whilst an interesting and useful contribution to the literature, included three studies that do not appear to be concerned with washback as defined. Of these three, two were experimental in design (Khoshsima, Saed & Mousaei, 2018; Khodabakhshzadeh, Zardkanloo & Alipoor, 2017) and one involved action research aimed at determining what might constitute effective test-preparation (Munoz, 2017). Unlike experimental and action research studies, however, washback is not an intervention, but is what teachers and learners are observed

to do in response to a test. This is not always what we might recommend them to do, and it is helpful to preserve a distinction between the actual, the desirable, and the experimental.

Studies of effective test-preparation strategies are of interest in their own right, and may help to inform both test development and programmed implementation, but they are not studies of washback per se. An understanding of the processes which operate alongside washback may help researchers to more clearly contextualise research and avoid blurring the established concept of washback.

#### **7. Conclusion**

The present paper has sought to bring together strands from different sources to present a fresh way of looking at the role that tests might play in driving educational change by contextualising the processes involved and how they relate to each other. Although the ideas presented are not new, at least one of them has been neglected for decades and it is suggested that without wash-forward the bigger picture is incomplete. The demands of a new test may provide the motivation for change, but interventions such as training programmes, and appropriate textbooks (i.e. programmed implementation) have an important role to play in increasing the potential for high-stakes tests to generate positive impact. Failing to equip teachers and learners with the knowledge and skills required of them may mean that they have little option but to continue doing what they know, and this may even lead to the test being forced to change (wash-forward).

The visualisation presented in figure 1 may be of use to students and researchers seeking to understand washback and related ideas, but it is perhaps of greater potential value in communicating with non-specialists

in assessment, such as policy makers who have the power to make decisions about new tests and their implementation. Without over-simplifying ideas, it is essential that the assessment community is able to communicate other educational stakeholder groups in ways that non-specialists can understand and integrate with their existing knowledge. It is hoped the present paper is a helpful contribution to that effort.

### Acknowledgements

My thanks to Dianne Wall for her insight and help while supervising the dissertation that sparked the thinking outlined in this paper, and to my colleagues at the British Council, especially Barry O'Sullivan, Robin Skipsey, Chie Yasuda, Mina Patel, Sheryl Cooke, Richard Spiby, and Jamie Dunlea, for their comments and questions in discussion as the idea for the paper took shape. I would also like to thank the anonymous reviewer whose insightful comments were very helpful in improving the final draft. Any remaining deficiencies are entirely my own.

### References

- Ahmed & Rahman, (2019). Review of empirical studies of washback in language testing. *Bangladesh Maritime Journal*, 3(1), 150-162.
- Alderson, J.C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: a study of washback. *Language Testing*, 13, 280-297.
- Buck G (1988). Testing listening comprehension in Japanese university entrance examinations. *JALT Journal*, 10, 15-42.
- Cheng, L. (2002). The washback effect on classroom teaching of changes in public examinations. In Savignon, S.J. (Ed) *Interpreting communicative language teaching: concepts and concerns in teacher education*. Yale University Press
- Cheng, L. (2005). *Changing Language Teaching Through Language Testing: A Washback Study*. Cambridge, U.K: Cambridge University Press.
- Henrichsen, L. (1989). *Diffusion of Innovations in English Language Teaching: The ELEC Effort in Japan, 1956-1968*. Greenwood Press Inc., Connecticut, USA.
- Hughes, A. (2003). *Testing for language teachers* (2<sup>nd</sup> ed.). Cambridge, U.K.: Cambridge University Press.
- Johnstone, R. (1994). Research on language learning and teaching: 1993. *Language Teaching*, 27(3), 145-156. doi:10.1017/S0261444800007795
- Khodabakhshzadeh, H., Zardkanloo, R., and Alipoor, I. (2017). The effect of mock tests on Iranian EFL learners' test scores. *International Journal of Education & Literacy Studies*, 5(3), 47-51.
- Khoshsima, H., Saed, A., and Mousaei, F. (2018). Exploring the effect of teaching test-taking strategies on intermediate level learners on reading section of IELTS; learners' attitude in focus. *Advances in Language and Literary Studies*, 9(2), 4-9.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13, 241-56.
- MEXT (2014). *Report on the future improvement and enhancement of English education (outline): Five recommendations on the English education reform plan responding to the rapid globalization*. Retrieved from <http://www.mext.go.jp/en/news/topics/detail/1372625.htm>
- Munoz, R. E. (2017). The effect of washback on EFL learners' attitudes toward tests. *Studies in English Language Teaching*, 5(3), 516-530.
- O'Sullivan, B. (in press). The comprehensive learning system. *British Council perspectives on English language policy and education*.
- Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing*, 22 (2), 142-173.
- Roberts-Gray, C., & Gray, T. (1983). Implementing Innovations. *Knowledge: Creation, Diffusion, Utilisation*, 5(2), 213-232.
- van Lier, L. (1989). Reeling, writhing, drawing, stretching, and fainting in coils: oral proficiency interviews as conversation. *TESOL Quarterly*, 23(3), (Sep. 1989), 489-508.
- Wall, D., & Alderson, J.C. (1993) Examining washback: the Sri Lankan impact study. *Language Testing*, 10, 41-69.
- Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing*, 13, 318-33.

# KIỂM TRA – ĐỘNG LỰC THAY ĐỔI TRONG GIÁO DỤC: XÁC ĐỊNH TÌNH HÌNH TÁC ĐỘNG ĐỘI NGƯỢC VÀ KHẢ NĂNG TÁC ĐỘNG THỨC ĐẨY

Gordon Allan

*Hội đồng Anh, 1-2 Kaguruzaka, Shinjuku-ku, Tokyo, 162-0825, Nhật Bản*

**Tóm tắt:** Mặc dù mục đích chính của các bài kiểm tra là đo lường năng lực thí sinh, nhưng không có gì lạ khi các bài kiểm tra được triển khai trong các hệ thống giáo dục với mục đích, ít nhất là một phần, để thúc đẩy sự thay đổi trong thực tiễn giáo dục bằng cách đưa ra yêu cầu mà giáo viên và người học được hi vọng có thể đáp ứng. Tuy nhiên tác động của kiểm tra đánh giá lên hoạt động dạy và học (tác động đội ngược - washback) không phải là điều duy nhất có thể xảy ra và cũng thường không xảy ra đúng như mong muốn. Nghiên cứu này thu thập các ý tưởng khác nhau nhằm đưa ra các khả năng khác. Mô hình đưa ra liên quan tới các khái niệm “điều chỉnh bài thi” (adaptive implementation) và “điều chỉnh hệ thống” (programmed implementation) được lấy từ mô hình của Henrichsen về các điều chỉnh mang tính đổi mới trong hệ thống giáo dục. Tác động của kiểm tra đánh giá lên hoạt động dạy và học (washback) được cho thấy diễn ra song song nhưng cũng riêng biệt với điều chỉnh hệ thống (programmed implementation). Bức tranh toàn cảnh được hoàn thiện với khái niệm của Lier về “wash-forward” – tác động của hoạt động dạy và học lên kiểm tra đánh giá, một khái niệm được đưa ra lần đầu tiên vào năm 1989 nhưng sau đó đã dần bị lãng quên. Tác động thúc đẩy hoạt động dạy và học lên kiểm tra đánh giá (wash-forward) được minh họa bằng một ví dụ từ việc triển khai Bài thi tiếng Anh quốc gia National Matriculation English Test (NMET) tại Trung Quốc. Ví dụ này nhằm đưa ra một góc nhìn dễ hình dung, mới mẻ, và đầy đủ hơn về các quá trình giới thiệu một bài thi mới trong chiến lược thúc đẩy những thay đổi trong thực tiễn dạy và học; một ví dụ rất phù hợp với phong trào hướng tới kiểm tra đánh giá bốn kỹ năng tiếng Anh hiện nay ở Đông Á và trên toàn thế giới.

**Từ khóa:** Tác động của kiểm tra đánh giá lên dạy và học, tác động của dạy và học lên kiểm tra đánh giá, điều chỉnh hệ thống, đổi mới giáo dục, cải cách thi cử