Washback in Education:

A Critical Review and Its Implications for Language Teachers

Ka Long Roy Chan (roylongckl@gmail.com)

The Chinese University of Hong Kong, Hong Kong

http://orcid.org/0000-0002-4521-9124

Abstract: Even though washback has been widely researched in recent years, especially on its form in various intra- and inter-national examinations (e.g. IELTS and TOEFL), research on how washback affects teachers and their teaching practices is still scarce. The aim of the current paper is to provide an up-to-date and holistic review on the theories in washback as well as to draw researchers' and TESOL teachers' attention to how washback should be examined with relation to language education, especially in teacher trainings. Key models in washback and several recent studies which investigate washback in different aspects are discussed, and at the end of the paper, implications and suggestions on washback for teachers and researchers are addressed. More should be done on washback especially on how teachers should respond to the effect.

Keywords: Washback, Teacher Training; Language Testing

Introduction

Washback (or 'Backwash') is a widely discussed topic in fields like applied linguistics and general education. It refers to "the influence of testing on teaching and learning" (Barnes, 2017, p.1). In general, researchers agreed that washback is a sophisticated mechanism which is affected by a multitude of factors. For example, direction of washback (positive or negative) (Watanabe, 2004), washback intensity (Green, 2007), different stakeholders in washback (Pan, 2009) etc. Most of the previous studies focused on examining whether certain examinations create a positive or negative washback (Kılıçkaya, 2016). Meanwhile, the subjects in the

previous studies on washback were most likely students. Given that "washback means the degree of behavioral change for *teachers* and *learners* resulting from the introduction and use of tests" (Park, 2018, p.125, italics by the author), and that teachers play an irreplaceable role in education, there is certainly a strong need for more emphasis to be placed on the washback on teachers.

Moreover, there exists a number of models and different foci in studying washback. While there are studies on washback which focus on specific places like Turkey (Toksöz & Kılıçkaya, 2017), and specific tests like IELTS (Allen, 2016; Green 2007), TOEFL iBT (Barnes, 2016, 2017) etc., there is seldom an attempt to provide a holistic review on all these models and on all existing studies. Furthermore, research on how washback impacts teachers and their teaching practices is seldom found in existing literature. Since washback is an effect exerted among language testing, language teaching, and language learning, and given the lack of research on its impact on teaching , it is therefore needed to call for attention to this particular part of study, especially when teachers are important in a washback as they are the practitioners of the teaching (Chan, 2018). The study of washback is important to teachers' training because it could be helpful in various aspects of teaching, for example, the evaluation of the teaching materials (Bernes, 2016) and the adjustment in achieving an effective teaching (Cholis & Rizqi, 2018). The aims of the current review paper are two-fold. First, it hopes to provide a holistic review, which is up-to-date, on several critical models and research directions of washback in order to draw a general view for teachers and researchers in the field. Second, it strives to draw researchers' attention to the washback on language teachers especially when the existing literature on how washback affects teachers is scarce. In the following sections, the types and the models of washback studies (for example, the terminology and directions of washback) will be reviewed first, followed by an introduction of different recent washback studies which are categorized based on their focus of investigation. This paper ends with the implications and suggestions of washback study, especially on teachers' trainings, which are generated from previous studies.

Types of Washback and Models of Washback

To begin with, a definition on the terminology of washback is needed. Even though the terms *Washback* and *Backwash* are often used interchangeably among researchers (Wang & Bao, 2013), Green (2007) commented that Backwash is commonly annotated with negative meaning within applied linguistics and Washback "has gained in currency and is now generally accepted in the applied linguistics literature" (p.2). Washback will therefore be used throughout this paper.

Washback can be generally defined as "a neutral concept meaning influence (which)… are the things that teachers and students do because of the existence of tests" (Park, 2018, p.125). The effects from washback can be divided into the effects on learning (to learners) and on teaching (to teachers) (Green, 2007). Studies on washback can be divided into the effects on a micro-level and a macro-level (Bachman & Palmer, 1996). On a micro-level, it affects the language learning and teaching in a classroom context (Hakim, 2018); on a macro-level, it influences the entire education system, ranging from curriculum design to administration (Chan, 2018). Despite all these categorizations, washback is most often studied within the positive-negative dichotomy in which different stakeholders are involved.

The effects of washback are often investigated in relation to whom the effects are exerted on. Hughes (1989) and Bailey (1996) had mapped the potential stakeholders who experience the effect of Washback in their *Trichotomy Backwash Model* (*Table 1*). The model shows the stakeholders in Washback (teachers, students, policy makers, etc.), the actions that are taken by the stakeholders (learning and teaching process), and the products that are involved (the test).

Table 1. Trichotomy Backwash Model (Bailey, 1996, cited in Chan, 2018)

Different Parts in Backwash	Example
1. Participants	People like students, teachers, material designers, and policy- makers whose attitudes might potentially be influenced by the examinations.
2. Processes	The actions taken by the participants during the learning and teaching process concerning the examinations.
3. Products	The actual output of teaching or the results of the examinations.

Hughes (1989) described that the washback of a test can be either positive or negative depending on whether the effects from the test is beneficial or harmful to learners and teachers. Pan (2009) summarized both positive washback and negative washback on both micro- and macro-level washback to learners and teachers and below is a modified version (Table 2).

Table 2. Examples of Positive and Negative Washback in Different Levels (Modified from Pan, 2009)

		Positive Washback	**Negative Washback**
Micro-level	Students	are motivated to work harder	learn only knowledge that is tested; have a negative attitude towards learning; learning motivation is lowered
	Teachers	cover the subject more thoroughly and complete the syllabus within a time limit	narrow the curriculum and only cover those tested subjects
	Others	teaching-learning process is encouraged	Anxiety is created for both learners and teachers
Macro-level	Education System	The authority will review and introduce new textbooks and curricula to achieve the goals of teaching and learning which fit the expectations from society	The authority uses tests to promote political agendas and seizes control over the educational system

In short, positive washback results in a motivated learning and teaching environment where learners acquire what is essential with joy and teachers transmit

knowledge to students without much pressure whereas negative washback leads to a distorted learning environment in that all the participants in the washback suffer from the test.

Based on the positive-negative dichotomy, Green (2007) created a model to explain how positive or negative washback is generated from a test. Green (2007) believed that there are two elements in a test: the *focal construct* and *the test characteristics*. The former is what the curriculum designers desire in delivering (or what the test originally wants to provide to students) and the latter is what is exactly assessed in the test, including the content, test format, complexity of tasks, etc. Positive washback is achieved when the two ideas are overlapped to a great extent; in other words, the test "reflects the focal construct as understood by the course provider and learners" (p.14); negative washback is created if there is only little (or no) overlapping on the two ideas. Below is an illustration of this direction of washback adopted from Green (2007) (*Figure 1*):
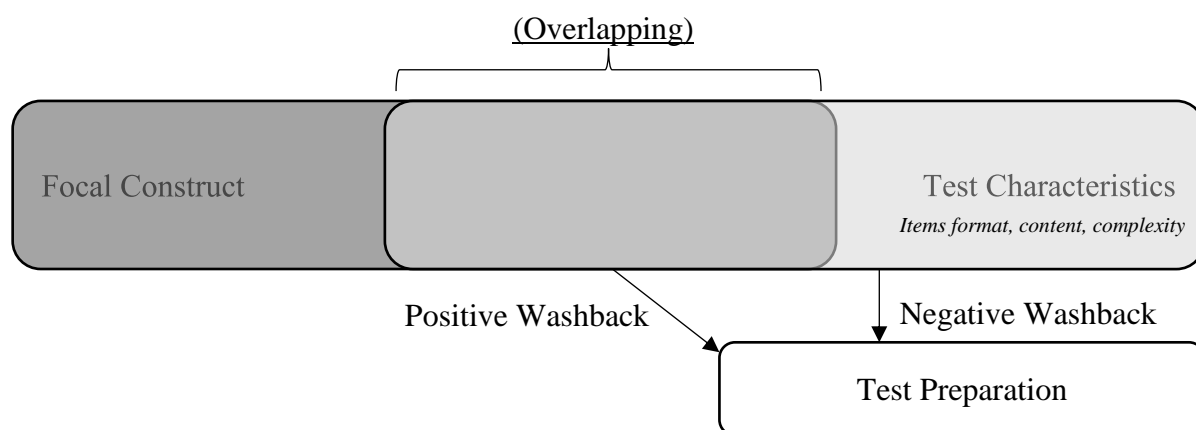

Figure 1. Direction of Washback (Adopted from Green, 2007)

Moreover, there is a scale of intensity of washback. Toksöz & Kılıçkaya (2017) mentioned that, in a general sense, "if a test is significant for the test taker, it displays strong washback; yet, if it is not fundamental at all, it presents weak washback" (p.185). Watanabe (2004) described the intensity of washback as how it affected the classroom context:

If the test has a strong effect, then it will determine everything that happens in the classroom, and lead all teachers to teach in the same way toward the exams. On the other hand, if a test has a weak effect, then it will affect only a part of the class- room events, or only some teachers and students, but not others. (p.20)

Green (2007) aligns the importance of washback to the difficulty of test and the intensity of washback; he concluded that strong washback effect will be expected when:

1. *the success in the test is more important than skill development;*
2. *the test is challenging and*
3. *all the participants (learners, teachers, developers) hold the same view on the test* (Green, 2007)

Combining Hughes' (1989), Bailey's (1996), Watanabe's (2004), and Green's (2007) views, it can be said that washback is the influence from a test on teaching, learning, and education planning in which a number of participants (teachers, learners, planners) are involved. The effect can be positive or negative with different intensity based on how significant the test concerned is and the effect will be intensified when the items that are tested are different from the initial design expected.

While the above are most related to what create washback (or the factors of washback) and how washback is categorized, Chan (2018) described how washback works in education context. Based on the Trichotomy Model, Chan (2018) purposed a Washback Cycle for Teachers (*Figure 2*).
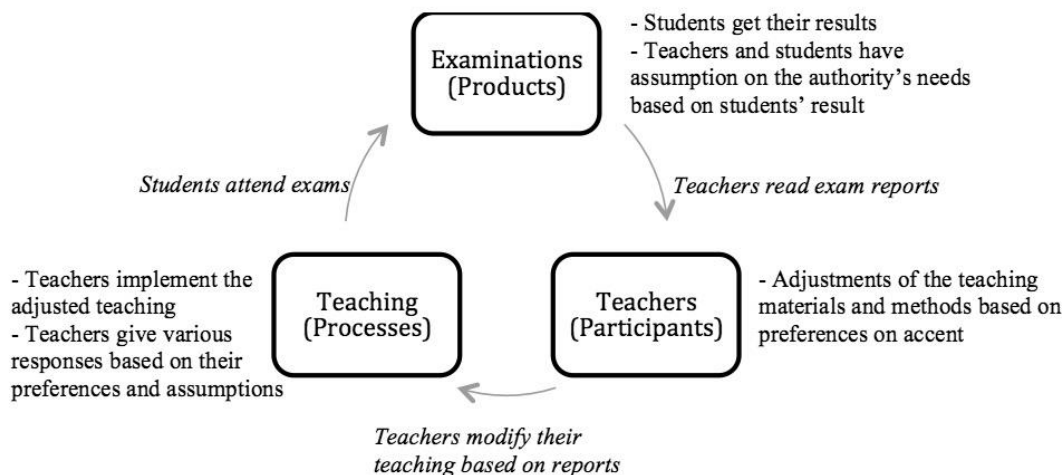
Figure 2. Washback Cycle for Teachers (Adopted from Chan, 2018)

This model is in a cycle because it indicates that "washback seems to be an ever-going process and the teachers would modify their teaching year by year (because of the washback)" (Chan, 2018, p.37). The examination is at the top which indicates it is the test which initiates the cycle. The result of the exam affects how teachers teach. The teachers will then adjust their teaching accordingly to fit the examination (teaching process) and the students who receive the modified teaching (learning) take the examination (product) which will generate another round of result for teachers to modify their teaching again.

In summary, different types of washback (for example, positive and negative), the stakeholders in washback and how washback works are discussed in this section. In the next part, some major recent washback studies with different foci will be introduced.

Recent Studies on Washback

Even though washback is categorized differently as mentioned in the previous section, studies on washback are seldom done under those categorizations. It is because most of the time washback is tested as an effect from an examination; therefore, the foci of these studies are often on the tests concerned. Intra- or International high-stakes examinations are therefore frequently the subjects

concerned in these studies. Also, within these studies, researchers normally emphasize the effects exerted on the participants (leaners, teachers, policy makers, etc.). Therefore, in this section, several recent studies on washback, especially with respect to teachers' performance and education, will be selected and introduced under the categorization of *washback in high-stakes examination* and *washback to learners and teachers*, which will lay the ground for the suggestions to teachers and researchers in the next section.

Washback in Intra- and International High-stakes Examination

Since "high-stakes tests have often been used in language education to change teaching and learning practices as intended by policymakers and test designers" (Barnes, 2017, p.2), washback studies about high-stakes examinations are frequently done to see whether the test achieves certain purposes and whether the test brings positive or negative washback to students or teachers. These high-stakes examinations include intra-national examinations (e.g. state-wide university entrance examinations) and international examinations (e.g. IELTS or TOEFL). It is very important for teachers to understand how washback is portrayed in these examinations as this may help them adjust their teaching.

Cheng (2004) and Tsang (2017) investigated two generations of English high-stakes examination in Hong Kong, the Hong Kong Certificate of Education Examination (HKCEE) and the Hong Kong Diploma of Secondary Education Examination (HKDSE). By analyzing the data from questionnaires which sampled over eight-hundred students sitting for HKCEE, Cheng (2004) concluded that the examination has an effect on changing students' preferences on the choice of learning activities, and that students may choose to take part in activities which are highly relevant to the tested items. Tsang (2017) examined the washback effects of HKDSE (pre-university examination in Hong Kong which replaced HKCEE in 2011). The analysis of three focus group interviews with twelve secondary six students and one-hundred-fifty questionnaires in Tsang (2017) revealed that there are multiple washback effects brought by the examination to the students. The effects included

the needs to join intensive exam drilling, extra tutorial courses, etc. Also, Tsang (2017) identified the mediating factors of the students to take the exam such as influence from teachers, parents, peers, and society.

Pan (2013) investigated washback effect from the exit examinations in Taiwan universities. Employing a mixed-method approach, Pan (2013) analyzed data from one-hundred-sixty questionnaires, twenty-five interviews, and fifty classroom observations and drew a conclusion that the exit examinations of university created effects on the way of teaching and learning, in which teachers concerned more on tested items. Toksöz and Kılıçkaya (2017) summarized the findings of fifteen washback studies in Turkey and concluded that various national examinations in Turkey have a negative washback on both teachers and students, especially when students overemphasize the study of some tested skills and teachers face a narrowing of curriculum which impacts their teaching.

For international examinations, Allen (2016) investigated washback of IELTS in a Japanese university, and one-hundred-ninety questionnaires and nineteen interviews were conducted to elicit the data. He discovered a positive washback in the test as students strategically shifted their focus of learning towards different English skills that are concerned in the test. Green (2007) examined the learning performance of over four-hundred university students on IELTS writing and questioned the use of preparatory courses of EAP courses. The results of Green (2007) suggested that a test-focused course may not improve test performance of students. Barnes (2016, 2017) examined the washback of another international examination, TOEFL iBT. By analyzing data from class observations and interviews of teachers, Barnes (2017) compared how teachers of general English and teachers of international examination differ in teaching and concluded that "teaching practices differed depending on the teaching context, or more specifically, the role and purpose of the course" (p.1). Barnes (2016) investigated how the selections and the use of teaching material creates influences in the TOEFL iBT examination in Vietnam. By analyzing the classroom materials and classroom observation of four language

teachers in Vietnam, Barnes (2016) suggested that test materials are keys to the teaching of such kind. Also, the test not only affects the contents of teaching, but also the pedagogies used by teachers; for example, it affects how the teachers plan their lessons in terms of time and material allocation, which leads to a conclusion that "(the test and the test materials) have the ability to influence teacher behavior and therefore how they teach" (p.171).

In short, it is clear that different research studies have been done on various levels mentioned previously. For high-stakes examinations, most of the time the concern is whether the washback from examinations is a positive or negative one. However, it is normal to discover that scholars find it hard to determine whether a test is generally positive or negative as there is a multitude of factors governing the examinations, for example, the different stakeholders in washback that is introduced previously. Moreover, many of these washback studies failed to address the problems from teachers' and students' perspectives as they focused more on the tests rather than the teachers and students. Among all the stakeholders, two of them – learners and teachers – are the most directly associated with washback and there have been several studies on examining the washback on them.

Washback to Learners and Teachers

As the above studies suggested, washback may potentially create huge effects on teaching and learning as what Chan (2018) described that "washback is an important effect in language testing as it alters how the teachers teach and eventually how the students learn" (p.38). Even though the current paper mainly focuses on washback with respect to teachers, how it affects students should also be discussed as teaching and learning are inseparable and teachers play a key role in these processes. Therefore, here in this part, several recent research studies on how washback affect learners and teachers will be shown.

To begin with, Park (2018) conducted a research on how washback affects students in preparing for an English exam (CSAT) in Korea. One-hundred Korean students

were surveyed and Park (2018) reported that the students all experienced a negative washback in their learning process. Students were directly impacted by the examination on their choices of activities inside as well as outside the classroom. The results also showed that students may choose to focus on the tested features rather than the non-tested features even if they are more interested in learning the latter ones. Park's (2018) result is in line with Allen's (2016) and Tsang's (2017) that examination brings washback to students' preferences on learning strategies and it can be considered as a negative washback. Tsang (2017) summarized four main types of washback from the examinations for students:

1. *Informal ways of training the preferred section outside classroom;*
2. *Selective attention in English Language learning;*
3. *Intensive paper-and-pencil drills on the preferred section;*
4. *Enrolment in section-focused tutorial classes.* (Tsang, 2017, p.60)

These can be seen as the major effects that are exerted on students from examination as these all seem to narrow students' interest in learning holistically. Students are forced, intentionally or unintentionally, to care more about the tested features by taking intensive drills and extra courses. This can be confirmed by the result of Saglam and Farhady (2019) - they conducted their survey and analysed how washback affected 147 students in an English proficiency exam in Turkey. The results from interviews and the scores of a pre- and post-test echo with Park's (2018) that a negative washback was found because students started to adopt a more test-oriented mindset, which is not ideal for language learning. Also, students found that there may be a mismatch between the learning materials and the actual exam. However, positive washback was also found among the students who discovered that the skill sets are transferable from one paper to another.

Similar effects were examined among teachers. Chan (2018) investigated how pronunciation teaching is affected among teachers in Hong Kong, in particular the

pronunciation of two varieties of English used in Hong Kong (Hong Kong English[1] and British English). By analysing data from one-hundred-eighty questionnaires and the interviews with twenty-eight in-service teachers, Chan (2018) concluded that negative washback exists in the pronunciation teaching in Hong Kong as teachers expected a standard pronunciation (British and/or American English) from the examination which does not appear to be stated in the curricula and students are prompted by their teachers' teaching to follow the standard pronunciation which does not exist. It is reported as the *Washback Cycle for Teachers* in which how the teachers teach, how students learn, and how the exam is tested are all interlinked and it runs in a circle. The results of Chan (2018) echo with Barnes' (2017) that the teachers' belief may bring influences on their teaching practices and directly affect students' learning, which initiates the washback effect.

These studies show how students and learners are affected by washback in exams; however, it is also worth noticing that these always fail to provide a more generalized picture for broad education because they focus only on one specific exam; it is also one of the most common limitations of current washback studies. To summarize this part, it can be seen from the previous research that washback can potentially affect both students and teachers in terms of their learning and teaching process. More importantly, the effect is not linear but cyclic that may intensify over time when one affects the other, and it goes on in the Washback cycle. Different stakeholders in the washback (in the Trichotomy of Washback) should actually be alerted. Among all the stakeholders, teachers should be the foremost to pay attention to as they are the practitioners of teaching and therefore in the last part of this paper, there are some implications of these washback studies for teachers, as well as for TESOL researchers.

---

[1] Hong Kong English is a new variety of English under the world Englishes paradigm; more can be found in Chan (2016/2017) and Hansen Edwards (2019).

Implications to Teachers, Researchers and Concluding Remarks

As mentioned earlier in the introduction, even though the number of studies on washback has been increasing over the past two decades, there has not been sufficient attention on how washback impacts teaching practices. Moreover, since previous studies were done separately on different stakeholders in washback, there is a lack of suggestions for teachers to understand how they could benefit from these studies, for example, getting insights into how to modify their teaching in order to promote a positive washback. Therefore, in this final part, how teachers could benefit from washback research and how researchers should further investigate washback in teachers' education context will be examined. In a very general sense, a positive washback should be promoted as it creates a more motivational learning environment for learners. Hakim (2018) expressed that teachers are the keys to promote positive washback on the micro-level in the classrooms as they have controls over the teaching materials and students' behaviors; more importantly, teachers are the bridges between students and policy-makers and the test designers, who are stakeholders which affect the washback on the macro-level. Teachers are therefore important in helping achieve positive washback. In order to help teachers in promoting positive washback, Spratt (2005) suggested several ways:

1. *Ensure that educational goals are pursued in the classroom;*
2. *Use of both authentic tasks and authentic texts in testing;*
3. *Provide full feedback to students on class tests.* (Spratt, 2005, p.25-26)

In Spratt's (2005) suggestions, the above three are especially relevant and feasible in the real education context. Setting a valid objective for students related to the test is important as it ensures that students follow closely to the requirements of examination. At the same time, since teachers have controls over the classes, they should use authentic texts and tasks alongside the tested items to make sure that students acquire knowledge to handle the needs in examinations as well as in real life. At last, teachers should provide full feedback to students based on the examination reports from the authority. Acting as the bridge between students and

the examination authority, it is the teachers' responsibility to let the students know what exactly the authority wants.

For researchers, research in washback provides important insights into the whole education process including teaching, learning, and assessment. However, as Allen (2016) admits, there are several limitations among all washback studies that need to be tackled in future studies – i) the use of self-report data is often too subjective that cannot reflect the real situation; ii) the use of indirect measurements (test scores) create problems as there is not a necessary relation between direction of washback and student performance; iii) students who are surveyed are always high-achievers which do not represent a whole picture of learners. Researchers should consider in the future studying washback in ways that can overcome these default limitations. Moreover, previous studies often focused on a single test and a single group of people in a washback without acknowledging that washback is the interaction among various stakeholders in the education system. As Barnes (2016) remarked at the end of his study that " (the study) lacks student and teacher perspectives and attitudes towards teaching, learning and testing", she concluded that more should be done on washback, including research which incorporates different methods of data collection (e.g. interviewing and observing teachers), which may eventually allow a better understanding on the issue. More should be done on investigating how these stakeholders are interacting with each other, which creates different washback, not in one specific exam but in a broad education system, which would potentially provide a more generalized picture to the language teachers in general. At last, studies should be done on how washback works in cycle among the stakeholders and how different parties are interlinked.

To conclude, the current review article summarized models that are related to washback - the *Trichotomy Model* and the *Washback Cycle for Teachers*– which describe the stakeholders in a washback and how washback works in teacher's perspectives. The dimension of washback has been discussed and relevant studies conducted recently have been introduced. It is important to know that "washback is an

important effect in language testing as it alters how the teachers teach and eventually how the students learn. Research of this kind is therefore necessary in providing insights into whether a specific test is serving its aims" (Chan, 2018, p.38). It is therefore hoped that this article could provide insights for TESOL teachers and researchers in how washback may have influenced the teaching and learning processes.

References

Allen, D. (2016). Investigating washback to the learner from the IELTS test in the Japanese tertiary context. *Language Testing in Asia, 6*(7), doi: 10.1186/s40468-016-0030-z

Bachman, F. L. & Palmer, S. A. (1996). *Language testing in practice: Designing and developing useful language tests.* Oxford: Oxford University Press.

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

Barner, M. (2016). The washback effects of the TOEFL iBT in Vietnam. *Australian Journal of Teacher Education, 41*(7), 158-174.

Barnes, M. (2017). Washback: Exploring what constitutes "good" teaching practices. *Journal of English for Academic Purposes*, 30, 1-12.

Chan, K. L. R. (2016/2017). Attitudes towards Hong Kong English: Native English Teachers and Local English Teachers. *Asian Journal of English Language Teaching , 26*, 85-110.

Chan, K. L. R. (2018). Washback in English pronunciation in Hong Kong: Hong Kong English or British English? In Chan, K. G. D. et al. (Comps.), *Motivation, identity and autonomy in foreign language education* (pp.27-40). Singapore: NUS Centre for Language Studies.

Cholis, H. W. N. & Rizqi, F. (2018).Senior High School English Teachers' Perceptions on a High-stakes Test (SBMPTN): A Washback Study. *International Journal of Education and Literacy Studies, 6*(3), 47-52.

Cheng, L. (2004). The washback effect of a public examination change on teachers' perceptions toward their classroom teaching. In L. Cheng, Y. Watanabe and A. Curtis (Eds.), *Washback in Language Testing: Research Contents and Methods* (pp. 147–170). Mahwah, NJ: Lawrence Erlbaum.

Green, A. (2007). Washback to learning outcomes: a comparative study of IELTS preparation and university professional language courses. *Assessment in Education, 14*(1), 75-97.

Hakim, L. N. (2018). Washback effect in language testing: What do we know and what is its effect? *Jurnal Forum Didaktik, 2*(1), 59-68.

Hansen Edwards, J. G. (2019). *The politics of English in Hong Kong: Attitudes, Identity and use*. London: Routledge.

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

Kılıçkaya, F. (2016). Washback effects of a high-stakes exam on lower secondary school English teachers' practices in the classroom. *Lublin Studies in Modern Languages and Literature, 40*(1), 116-134.

Pan, Y. C. (2009). Review of washback and its pedagogical implications. *VNU Journal of Science, Foreign Languages, 25*, 257–263.

Pan, Y. C. (2013). Does teaching to the test exist? A case study of teacher washback in Taiwan. *The Journal of Asia TEFL, 10*(4), 185–213.

Park, S. (2018). Reinterpretation of the concept of washback in language testing. *SNU Working Papers in English Linguistics and Language, 16*, 125-142.

Saglam, A. L. G. and Farhady, H. (2019). Can exams change how and what learners learn? Investigating the washback effect of a university English language proficiency test in the Turkish context. *Advances in Language and Literary Studies, 10*(1), 177-185.

Spratt, M. (2005). Washback and the classroom: the implications for teaching and learning of studies of washback from exams. *Language Teaching Research, 9*(1), 5–29.

Tsang, C. L. H. (2017). *Examining Washback on Learning from a Sociocultural Perspective: The Case of a Graded Approach to English Language Testing in Hong Kong* [Master Thesis]. University College London: British Council Hong Kong.

Toksöz, I. and Kılıçkaya, F. (2017). Review of Journal Articles on Washback in Language Testing in Turkey. *Lublin Studies in Modern Languages and Literature, 41*(2), 184-204.

Wang, Y. and Bao, J. (2013). Washback effect on second language acquisition test: A case study of college English entrance exam. *Proceedings of the 2013 International Academic Workshop on Social Science*, 864–868.

Watanabe, Y. (2004). Methodology in washback studies. In L. Cheng, Y. Watanabe and A. Curtis (Eds.), *Washback in language testing: Research contents and methods* (pp. 19–37). Mahwah, NJ: Lawrence Erlbaum.