



# Washback of English Language Testing on ELL Teaching and Learning: A Literature Review

Ling Wang<sup>1</sup>

<sup>1</sup> Department of Teaching and Learning, College of Education, Austin Peay State University, USA

## Abstract

*Washback refers to the influence of testing on language teaching and learning. It is a complex educational phenomenon prevailing in various academic contexts. Based on the theoretical frameworks of washback, extensive empirical research has been conducted on large-scale, high-stake, or standardized national and international examinations. This paper discusses conceptual models of washback and reviews representative empirical studies of washback of English language testing on ELL teaching and learning during the last three decades. The findings indicate coexistence of both positive and negative washback in teaching contents, teaching materials, teaching methods, student learning, teachers' feelings and attitudes, as well as students' feelings and attitudes. Future studies could investigate the test mechanisms at both micro and macro levels to mediate intended washback on ELL language teaching and learning while minimizing its negative effects.*

**Keywords:** Washback, Testing, ELL, Teaching, Learning

---

## Introduction

Testing plays a unique role in our education system. Various testing formats, such as standardized, multiple-choice testing or portfolio assessment, have a powerful influence on language teaching and learning. Madaus (1988) claimed that "It is testing, not the 'official' stated curriculum, that is increasingly determining what is taught, how it is taught, what is learned, and how it is learned" (p.83). Swain (1985) argued that teachers "will teach to a test: that is, if they know the content of a test and/or format of a test, they will teach their students accordingly" (p.43). In addition, public examinations have impact on the attitudes, behavior, and motivation of teachers, learners, and parents (Pearson, 1988, p.98). Examination scores for various educational and social purposes are used extensively, which have strengthened the influence of exams on teaching and learning, no matter in general education or language education.

The concept of exam influence in the field of English language testing and teaching has various labels. "Backwash", "washback" and "impact" are some of the best-known terms (Alderson & Wall, 1993; Hughes, 1989; Wall, 1997). "Washback" and "backwash" are often used interchangeably since "the difference in terminology has no semantic or pragmatic significance whatsoever" (Alderson & Wall, 1993, p.115). As an inherently interesting phenomenon to English language teachers, researchers, policymakers, and others in their instructional and educational activities, "washback" in teaching English as a second/foreign language and applied linguistic literature has been discussed for a longer time.

While the existence of washback is widely acknowledged, consistent conclusions about washback have not been drawn. Shohamy (1993) proposed that "while the connection between testing and learning is commonly made, it is not known whether it really exists and, if it does, what the nature of its effect is" (p.4). Alderson and Hamp-Lyons (1996) stated that "Much has been written about the influence of testing on English language teaching. To date, however, little empirical evidence is available to support the assertions of either positive or negative washback." (p. 281). Recent studies of ELL learners' perspective on washback showed both positive and negative influences on their learning (Reynolds, 2010). Furthermore, both negative and positive washback effect on English teaching materials have been reported (Azadi & Gholami, 2013; Lodhi et al., 2018).

This paper reviews theoretical frameworks of washback and representative empirical studies of washback in English language testing in the last three decades, exploring its impact on ELL teaching and learning with respect to teaching contents, teaching materials, teaching methods, student learning, as well as attitudes and feelings of English teachers and learners.

## English Language Testing Theories

English language testing has experienced four stages of development, each appearing in diverse historical backgrounds and for the needs of different language teaching.

### *Pre-scientific Testing Period*

Before English language testing found its scientific basis last century, it was just simple replication of English language teaching. During this period, language was taught with the grammar-translation approach since English language was treated as knowledge mainly consisting of phonetics, grammar and vocabulary. By requiring students to read and translate classic literatures in English, teachers emphasized grammar rules instructed in their native languages. Therefore, most of students were only good at English reading and writing, while incompetent in listening and speaking. Accordingly, the focus of English language testing was in grammar and vocabulary. The most common testing methods for English language learning were translation, composition, and reading. Carroll and Hall (1985) questioned these highly subjective testing methods and claimed them as major deficiencies because such approach is “the narrowness of the criteria of performance and the capriciousness of the marking which was predominantly of an uncontrolled subjective type”. The first stage of English language testing, as well as English language teaching, placed great emphasis on English language form and therefore was called “code-focused” testing system, rather than “message-focused” testing system (Li, 1997).

### *Psychometric-structuralist Testing Period*

During the World War II, a large number of language specialists were in high demand. With the development of English language teaching, the subjective language testing system could no longer satisfy the demands in new historical and educational situation. More valid and reliable testing methods were needed. Based on structural linguistics and psychometric way of teaching, a new testing - psychometric testing emerged. According to the structural linguistics, language can be divided into elements at four levels: phonological, lexical, syntactical and cultural, which are taught and tested through four skills: speaking, listening, reading and writing. According to the psychometric testing theory, discrete-point objective test formats are called “closed” item types. The most frequently adopted formats for English language testing included “multiple-choice items, sentences with blanks to fill in, and sentences to be translated in various ways” (Xu, 2004). As the beginning of scientific language testing, psychometric-structuralist testing increases the fairness of language testing and makes large-scale testing possible, which contributed to the development of English language testing. However, psychometric-structuralist testing ignores context, which is a crucial property of language. Due to its emphasis on English language form and structure rather than practical communicative need, this testing approach is still “code-focused” (Li, 1997).

### *Integrative Testing Period*

The third stage English language testing is integrative testing, which overcame the deficiency of psychometric-structuralist testing that broke English language proficiency into pieces while neglecting the context. Using the linguistic basis of unitary competence hypothesis, integrative testing adopts dictation and cloze to measure English language proficiency as a whole. This English language testing approach required test takers to demonstrate their ability to control more than one level of language, such as morphology and syntax, at the same time, or even two English language skills, for example, reading and writing (Xu, 2004). However, cloze and dictation are better in certain contexts compared to psychometric-structuralist testing methods, they still cannot present convincing evidence that candidates are able to read, write, speak or listen in English in real-life contexts.

### *Communicative Language Testing Period*

The development of English language teaching inspired English language teachers to pay increasing attention to the actual “use” of English language in real-life situations. As a result, the language testing system also called for new approaches to test the candidate’s ability to use language properly in real contexts. This led to communicative language testing characterized by:

- 1) authenticity that requires the tasks in the test to resemble real-life situations;
- 2) interaction that encourages the interaction between the candidate and the tasks;
- 3) unpredictability, i.e., the information gap between the candidate and the tasks; and
- 4) context, including linguistic context as well as the context of situation (Baker, 1989).

All these characteristics assess not only linguistic accuracy, but the competence of function in the target language (Morrow, 1979). Since various abilities are tested in the communicative competence, numerous English language

testing instruments have been employed, such as multiple-choice items, ask and answer to assess receptive skills of listening and reading, interview, oral presentation and composition for productive skills of speaking and writing. The strength of communicative testing lies in the fact that it takes different levels of factors, such as linguistic, sociolinguistic, pragmatic, and strategic factors into consideration. It evaluates the candidate's competence to use the target language in real-life context while predicting their performance in similar tasks.

In summary, the evolution of English language teaching influenced the development of the language testing system. However, they did not evolve at the same rate nor in the same direction. Usually, language testing advances far behind language teaching due to the influence of historical, social, and economical factors.

### Theoretical Framework of Washback

Washback is defined as the influence of testing on teaching and learning. However, researchers look into this issue with different points of view. Some explored the value and extent of washback. For example, washback was considered as “a consequence of high-stakes exams” (Alderson & Wall, 1993; Hamp-Lyons, 1997). Shohamy et al. (1996) perceived it “as the link between testing, teaching and learning”. Washback was also seen as “a potential instrument for educational reform” (Pearson, 1988). Moreover, Messick (1996) claimed that washback can make teachers and learners do things “they would not necessarily otherwise do because of the test.” Many studies discovered the dichotomic or trichotomic directions of washback. For instance, Bailey (1996) and Messick (1996) described washback “as being potentially positive (beneficial), negative (harmful) or neutral”. Andrews et al. (2002) and Qi (2004) divided washback into “intended and unintended”. Alderson & Wall (1993) concluded that a direct and linear relationship exists between the stakes of a test and the strength of washback, i.e., the higher the stakes of a test, the stronger its washback.

In this review, the phenomenon of washback should be understood as rather than what is taught and learned in English language classes determines what will be tested. It is high-stake English language tests that play a determinative role and have a great impact on various aspects of English teaching and learning. Washback can have positive or negative value (Watanabe, 2004). Positive value of washback usually refers to those desirable influences that can help to improve teaching and learning while negative washback are influences that are not desired by English language teachers and learners.

#### *Alderson & Wall's Washback Hypotheses*

Alderson and Wall (1993) published an article entitled “Does Washback Exist?” that is considered as the start of “washback research” and has a great influence on all major research reports and literature reviews in the field of washback in language testing. Based on the analysis of test “washback”, fifteen possible washback hypotheses related to factors that have various effects on different persons are developed, which help to “identify cases where washback might be thought to have occurred, and to see what, how and why it did or did not occur” (Alderson & Wall, 1993).

#### **Possible Washback Hypotheses (WHs)**

1. A test will influence teaching.

This is the WH at its most general. However, by implication:

2. A test will influence learning

Since it is possible to separate the content of teaching from the methodology:

3. A test will influence *what* teachers teach and

4. A test will influence *how* teachers teach and therefore by extension from 2) above:

5. A test will influence what learners learn and

6. A test will influence how learners learn

However, perhaps we need to be somewhat precise about teaching and learning, whence

7. A test will influence the *rate* and *sequence* of teaching and

8. A test will influence the rate and sequence of learning and the associated:

9. A test will influence the *degree* and *depth* of teaching

10. A test will influence the degree and depth of learning

If washback relates to attitudes as well as to behaviours, then:

11. A test will influence attitudes to content, method, etc. of teaching/learning

In the above, no consideration has been given to the nature of the test, or the uses to which scores will be put. It seems not unreasonable to hypothesize:

12. Tests that have important consequences will have washback, and conversely

13. Tests that do not have important consequences will have no washback.

It may be the case that:

14. Tests will have washback on all learners and teachers.

However, given what we know about differences among people, it is surely likely that:  
 15. Tests will have washback effects for some teachers and some learners, but not for others.  
 (Alderson & Wall, 1993, pp. 120-121)

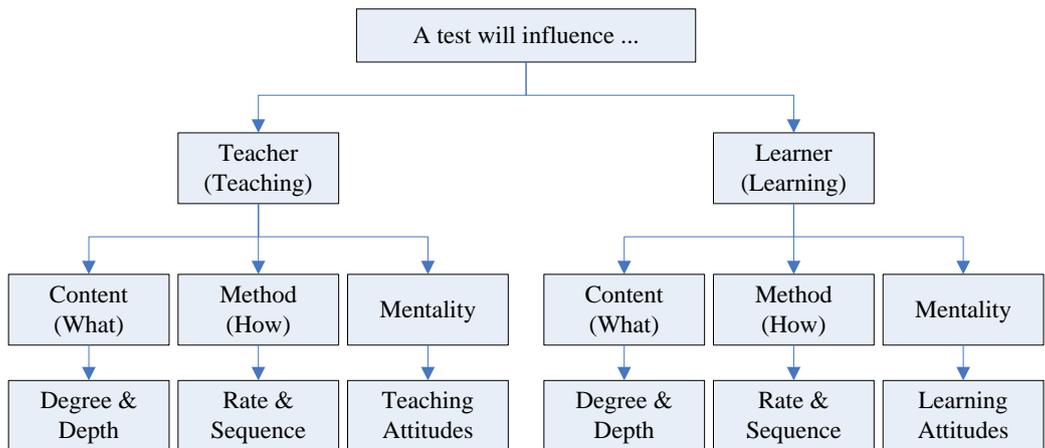


Figure 1. Illustration of Anderson and Wall’s Hypotheses (Shu, 2004)

Figure 1 shows that Alderson and Wall’s hypotheses are proposed from the dichotomic perspectives of teacher-learner and teaching-learning as well as the trichotomic levels of content, method, and mentality of both teachers and students (Shu, 2004). Their critical look at this phenomenon has outlined the territory for subsequent theoretical and empirical studies of washback in various contexts.

**The Trichotomy Model of Washback**

Hughes (1994) “made a distinction between washback on three constituents: the ‘participants’, the ‘processes’ and the ‘products’ of an educational system” (p.1). “Participants” refer to anyone whose perceptions and attitudes towards their work may be influenced by a test, such as classroom teachers or students, educational administrators, textbook developers, and publishers, etc. “Processes” are “any actions taken by the participants which may contribute to the process of learning”, such as the development of materials, the design of syllabus, changes in teaching methodology, and the use of test-taking strategies, etc. Finally, “product” refers to “what is learned and the quality of the learning” (Tsagari, 2007, p.10).

Based on Alderson and Wall’s Washback Hypotheses as well as Hughes’s distinction between participants, process and products, Bailey (1996, p. 264) proposed a model to delineate the complicated mechanisms of washback (see Figure 2). The impact of a test has two dimensions: 1) washback to learners, which refers to the direct impact of the test on test-takers, and 2) washback to the program, which means the impact on teachers, administrators, curriculum developers, and counsellors. Researchers and the participants are not only influenced by the test but also reciprocally have an impact on the test. This model no longer confines washback of a test solely to the micro aspects, such as teaching and learning. It also includes materials writers, curriculum designers, and researchers, focusing on the macro level of washback mechanisms.

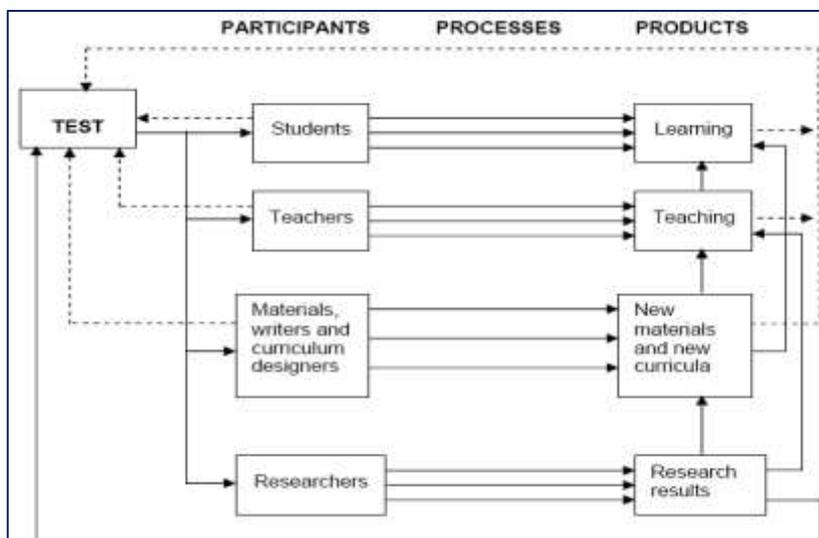


Figure 2. The Trichotomy Model of Washback (Bailey, 1996)

**Overt-Covert Washback Model**

Prodromou (1995) proposed a model divided into two categories: overt and covert washback. “Overt washback refers to the direct and evident teaching and learning to the test, for example, doing many past papers or mock exercises as preparation for examination, while, covert washback effect is deep-seated, often unconscious process.” Therefore, covert washback would result in “that teaching materials are becoming much more alike to the tests, and teaching procedures in the class are just like informal assessment” (Prodromou, 1995, p.15). It is easy to identify overt washback, while covert washback is more elusive and disturbing. Table 1 summarizes the characteristics of the communicative teaching and those of testing to contrast the typical teaching to the test with the ideal teaching. Qualities listed under “Testing” are symptoms related to either overt or covert washback.

Testing	Teaching
Exercises(multiple-choice)	tasks
Failure	success
Weakness	strength
Error phobia	learning from error
Product	Process
Marks	achievement
Fear	confidence
Anxiety	pleasure
Teacher control	learner independence
Textbook input	learner input
Judgment	support (from teacher and peer group)
Mistrust	rapport
Individualism, competition	the group, cooperation
Impersonality	personalization
Insensitivity	sensitive to the learners
Isolated sentences	text
Fragment of text	whole texts
Form	content
Culture-bound	culture-sensitive
Text-questions	lead-in, follow-up
Solemnity	humor
Boredom	interest
Extrinsic motivation	intrinsic motivation

Table 1. Overt-covert Washback Model (Prodromou, 1995)

**Intended Washback Model**

After the study of the intended washback of the National Matriculation English Test in China, Qi (2004) put forward a new model for the consequential aspect of validity, in which the intended washback was incorporated into the concept of validity (See Figure 3). Usually, washback refers to any influence caused by testing on teaching no matter it is intended or not. However, Qi (2004) suggested that washback should be divided into intended effects and unintended consequences, given that many studies revealed that tests have been used commonly as an agent for educational reform and the intended washback should be the focus of the consequential aspect of validity instead of unintended consequences.

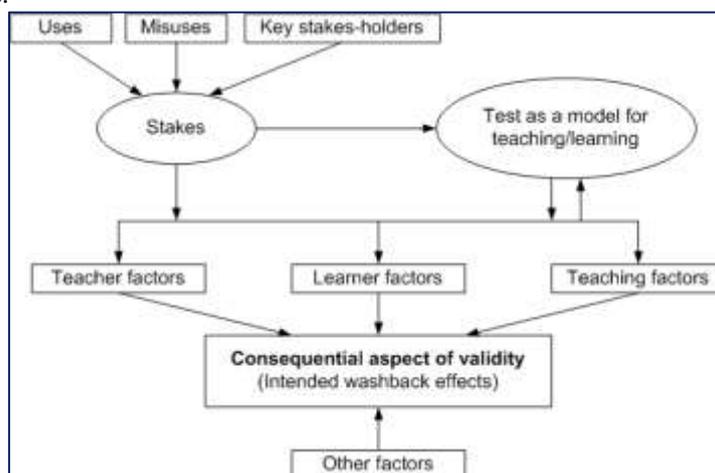
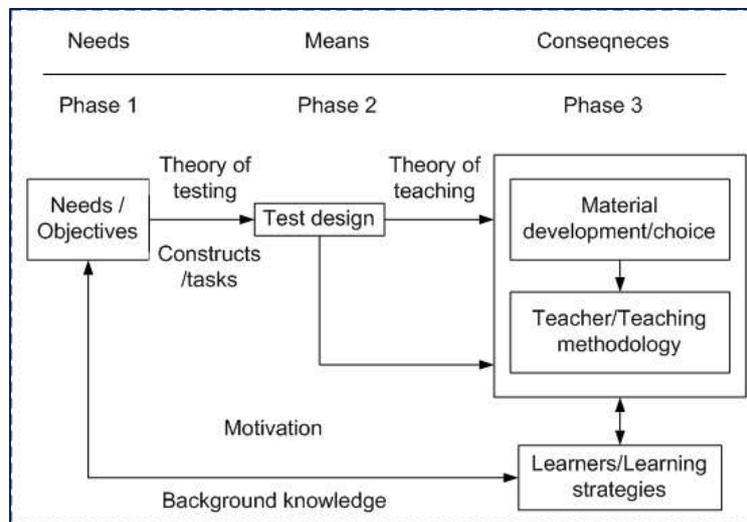


Figure 3. Intended Washback Model for the Consequential Aspect of Validity (Qi, 2004)

**Washback and Test Design**

Test design, as means of achieving intended washback, was not included in washback models developed in previous studies. Shahzad (2006), using the study of international teaching assistants, developed a conceptual washback framework that incorporates the needs and objectives of the educational setting and test design process (See Figure 4).



**Figure 4. An updated conceptual washback framework (Shahzad, 2006)**

In conclusion, these theoretical explorations looked into washback on English language teaching and learning from different perspectives, which laid important foundation for the empirical studies of washback in diverse contexts.

**Empirical Studies of Washback in English Language Testing**

With the theoretical hypotheses, models, and concepts, researchers have made empirical inquiry into washback of English language testing. Table 2 shows a summary of empirical studies in different contexts all over the world in the last three decades. The effects of washback on ELL teaching and learning are discussed from five aspects: teaching content, teaching materials, teaching methods, student learning, and attitudes and feelings of teachers and students.

Researchers	Year	Context	Test	Methodology
Hughes	1988	Turkey	University entrance examinations	<ul style="list-style-type: none"> <li>• Analysis of test scores</li> <li>• Teacher questionnaire</li> </ul>
Li	1990	China	Matriculation English Test (MET)	<ul style="list-style-type: none"> <li>• Teacher questionnaire</li> <li>• Local officer questionnaire</li> <li>• Student discussions</li> </ul>
Wall & Alderson	1993	Sri Lanka	O-level, English as an International Language (at the 11 <sup>th</sup> year of education)	<ul style="list-style-type: none"> <li>• Classroom observation</li> <li>• Teacher questionnaire</li> <li>• Teacher advisor questionnaire</li> <li>• Teacher interview</li> <li>• Student interview</li> <li>• Analysis of materials and tests</li> </ul>
Lam	1993 1994	Hong Kong	New Use of English (NUE) (end of secondary school)	<ul style="list-style-type: none"> <li>• Teachers questionnaire</li> <li>• Analysis of textbook</li> <li>• Analysis of test scripts and scores</li> </ul>
Shohamy	1993	Israel	Arabic as a Foreign Language Test English Foreign Language Oral test 3) L1 Reading test	<ul style="list-style-type: none"> <li>• Student questionnaire</li> <li>• Classroom observation</li> <li>• Interview</li> <li>• Analysis of document</li> </ul>

Watanabe	1996	Japan	English Language Exam for University Entrance	<ul style="list-style-type: none"> <li>• Classroom Observation</li> <li>• Questionnaire</li> <li>• Student interview</li> <li>• Teacher interview</li> </ul>
Alderson & Hamp-Lyons	1996	USA	TOEFL	<ul style="list-style-type: none"> <li>• Student interview</li> <li>• Teacher interview</li> <li>• Classroom observation</li> </ul>
Shohamy et al.	1996	Israel	1) Arabic Test 2) English Test	<ul style="list-style-type: none"> <li>• Student questionnaire</li> <li>• Interview</li> </ul>
Cheng	1997	Hong Kong	Revised Hong Kong Certificate of Education Examination (HKCEE)	<ul style="list-style-type: none"> <li>• Classroom observation</li> <li>• Teacher questionnaire</li> <li>• Student questionnaire</li> <li>• Interview</li> </ul>
Watanabe	1997	Japan	English Language Exam for University Entrance	<ul style="list-style-type: none"> <li>• Student interview</li> </ul>
Hamp-Lyons	1997	USA	TOEFL	<ul style="list-style-type: none"> <li>• Analysis of five TOEFL preparation textbooks</li> </ul>
Andrews et al.	2002	Hong Kong	Oral component of the Revised Use of English (RUE)	<ul style="list-style-type: none"> <li>• Videotapes of mock oral tests</li> <li>• Grading of oral tests</li> <li>• Discourse analysis</li> </ul>
Qi	2004	China	English Language Exam for University Entrance	<ul style="list-style-type: none"> <li>• Student interviews</li> <li>• Teacher interviews</li> <li>• Administrator interviews</li> </ul>
Shu	2004	China	English Language Exam for University Entrance	<ul style="list-style-type: none"> <li>• Questionnaire</li> <li>• Interview</li> <li>• Classroom observation</li> </ul>
Zhu	2006	China	High-stakes English Language tests	<ul style="list-style-type: none"> <li>• Questionnaire</li> <li>• Interviews</li> </ul>
Wang	2008	China	College English Test Level 4 (CET 4)	<ul style="list-style-type: none"> <li>• Analysis of paper</li> <li>• Questionnaire</li> <li>• Interview</li> <li>• Classroom observation</li> </ul>
Wang	2009	China	High-stakes English Language Tests	<ul style="list-style-type: none"> <li>• Questionnaire</li> <li>• Interview</li> <li>• Classroom observation</li> </ul>
Reynolds	2010	Australia	TOEFL	<ul style="list-style-type: none"> <li>• Survey</li> <li>• Interview</li> </ul>
Azadi & Gholami	2013	Iran	High school English Language Tests	<ul style="list-style-type: none"> <li>• Questionnaire</li> <li>• Classroom observation</li> </ul>
Adnan & Mahmood	2014	Pakistan	Higher Secondary School Certificate English exam	<ul style="list-style-type: none"> <li>• Teacher questionnaire</li> </ul>
Iyer	2015	Sri Lanka	English Language Tests at Universities	<ul style="list-style-type: none"> <li>• Questionnaire</li> <li>• Interview</li> <li>• Classroom observation</li> </ul>
Maniruzzaman	2016	Bangladesh	English Language Tests at Universities	<ul style="list-style-type: none"> <li>• Student Questionnaire</li> </ul>
Zou & Xu	2017	China	Test for English Majors for Grade Eight	<ul style="list-style-type: none"> <li>• Administrator Questionnaire</li> </ul>
Lodhi, et al.	2018	Pakistan	Secondary Level English Language Tests	<ul style="list-style-type: none"> <li>• Questionnaire</li> <li>• Test</li> <li>• Observation checklist</li> </ul>
Bokiev & Samad	2021	Malaysian	University English Language Assessment	<ul style="list-style-type: none"> <li>• Questionnaire</li> <li>• Interview</li> </ul>

**Table 2. Empirical Studies of Washback of English Language Testing**

### ***Washback on English Language Teaching Contents***

The reports of the washback effects are inconsistent in teaching content domain. Some studies show influence caused by exams, especially those new and revised exams. For example, Alderson and Wall (1993), in their Sri

Lanka study, stated that “the examination has had a demonstrable effect on the content of language lessons” (pp. 126-127). What will be taught was narrowed to those areas that will be most likely tested, such as writing and reading. Lam (1994) had a similar finding that those parts of the exam carrying the most marks were usually taught with emphasis. Likewise, Cheng (1997) discovered that teaching content changed accordingly with the introduction of a revised exam in Hong Kong, and Zhu (2006) revealed that teaching content often depended on what were covered in high-stakes English language tests in China. In a study of washback in College English Test (CET), a required national test for undergraduates in China, English language teachers perceived that teaching contents were tailored according to the outline of this exam, i.e., what they teach depend on what will be tested (Wang, 2008).

Different washback effects on teaching and learning curricula were also reported. For example, Shohamy et al. (1996), using questionnaires and interviews, report that the Arabic exam, as a low-stakes exam, had little effect on the teaching content while the high-stakes English as a Foreign Language had greater impact on the curriculum. Watanabe (1997, 2000) claimed, even though the exam contains the skills of listening or writing, teachers did not necessarily teach these skills. Class time distribution and class size are the factors related to the curriculum as mentioned in many studies. For instance, Lam (1994) found that exam classes were usually allocated more curriculum time. Shohamy et al. (1996) suggested that class time is not consistent for all examinations. Usually only high-stakes exams are given more class time. In their TOEFL preparation courses study, Alderson and Hamp-Lyons (1996) noted that some institutions allowed extra time in TOEFL classes while others did not. They also raised the factor of class size that may be affected by exams because more students were in exam classes than in “regular” classes. The results from questionnaires and classroom observations in a study of washback effects from English language testing for high school entrance examinations showed that the class time of the teaching and practice of the five English language skills, i.e., listening, speaking, reading, writing, vocabulary and grammar consistently reflect their weight distribution indicated in the exam study guide (Wang, 2009).

Zou and Xu (2017) conducted a study on the washback of Test for English Majors for Grade Eight (TEM8) in China with 724 English instructors and administrators. Their findings indicated that the course content and design of English courses in the universities were aligned with the requirements of TEM8. For example, syllabi for the writing and translation courses were designed to address the writing and translation part of the TEM8 test specifically. Lodhi et al. (2018) found that majority of English language teachers chose their teaching content based on English language test objectives rather than their students’ overall language learning needs, and they focused on the content that are relevant to the test to help their students succeed in the exam. Bokiev and Abd Samad (2021) discovered from their study on the Washback of an English Language Assessment System (ELAS) in a Malaysian University that English Language teachers had positive comments about washback on their course content: “ELAS had a facilitative impact on the teaching content as it helped them focus more on the development of skills that the programme was intended to develop.” (p.573)

In summary, washback on the teaching contents is not a phenomenon that can be explained in a simple way. Those studies indicate that, though it does not always work similarly in various situations, there is a tendency that washback on the curriculum is closely associated with the stakes of tests. The higher the stakes of tests, the stronger washback on the teaching contents.

### ***Washback on English Teaching Materials***

The effect of tests, especially high-stakes tests, on teaching materials drives the publication of exam-preparation materials, such as exam-oriented textbooks and past tests (Alderson & Hamp-Lyons, 1996). These materials are designed for students and teachers who prepare for such tests. Exam-related materials vary according to their emphasis. For example, some materials are designed to help test-takers get familiar with exam techniques, while other textbooks emphasize the development of relevant language skills. The studies reviewed mainly discuss washback on teaching materials in terms of their content and the use of materials.

A direct impact caused by high-stakes tests on the content of teaching materials is considered as evidence of washback by many researchers. For instance, Watanabe (1996) examined teaching materials which were adopted to prepare students for university entrance examinations in Japan. The materials “consisted of past exam papers and materials which were constructed by the instructors ... on the model of past exam papers”, which showed that “washback did exist on materials” (p. 325). Hamp-Lyons (1998) analyzed the content of exam preparation materials by a small-scale study of five TOEFL preparation textbooks. The findings revealed that “the skills promoted by the textbooks generally consist of (a) test-taking strategies and (b) mastery of language structures, lexis and discourse semantics that have been observed on previous TOEFLs”. Meanwhile, the books “relate quite exactly to the item types and item content found on the actual test rather than to any EFL/ESL curriculum or syllabus or to any model of language in use” (p. 332). Wang (1997) investigated teaching materials for the preparation of IELTS (International English Language Testing System) exams with a specially-designed instrument, the Instrument of Analysis of Textbook Materials (IATM). After studying sample units of two IELTS exam-preparation textbooks in detail, it was discovered that the IELTS test had an influence on the content and format of the preparation textbooks. Furthermore, it was noted that certain omission of the textbooks, such as

scoring profiles was a sign of negative washback since students would not be in a position “to monitor their own progress and where to put more effort when using these textbooks” because of such omissions (pp. 44-45).

Researches on the use of exam materials by teachers are mainly based on indirect research methods, such as teacher questionnaires and interviews. For example, Lam (1994), using teachers’ questionnaire, described teachers in Hong Kong as “textbook slaves” and “exam slaves” because a large number of teachers heavily relied on the exam textbooks as well as past papers in exam classes instead of using materials that aim “at maximizing students’ language learning”, and “they believe the best way to prepare students for exams is by doing past papers” (p. 99). Shohamy (1993, p. 15) also found from the three language tests examined that “many teaching activities became test-like, mostly as a result of the new textbooks, which were strongly influenced by the test.” Alderson and Hamp-Lyons (1996) reported that most teachers depended heavily on the use of exam materials, and their negative attitude towards the exam discouraged them from teaching creatively with their own materials. Cheng (1997), using teacher questionnaires and classroom observations, drew a conclusion that teachers’ adherence to the textbooks indicated washback on the content of teaching and it might be due to the fact that the textbooks in Hong Kong not only provide information and activities but also suggest teaching methods and time allocations. Wang (2009) observed that teaching to the test materials was dominantly used by 8th grade English language teachers in their English classes.

Azadi and Gholami (2013) showed “an overall negative washback effect of the high school English language tests on teaching materials” (p. 1340). Because English language tests did not cover listening, speaking, and writing, these important English language skills were not taught in class, which significantly narrowed down high school English course curriculum in Isfahan. The communicative competence was tested in the English language tests but reduced to only two sub-competencies: grammatical competence and textual competence. The scope of the tests highly restricted learning objectives and activities of the English class. Students committed minimum time on listening, writing, and speaking as these skills will not be tested. Instead they spent most of their time working on worksheets filled with grammar and translation exercises between Persian and English. The study recommended high school English language tests evaluate students’ practical communicative competence in English, which may generate positive washback on English teaching and learning.

Adnan and Mahmood (2014) studied the washback of Higher Secondary Certificate Examination (HSCE) on English language teachers. They reported that teachers prepared their teaching contents according to the test objectives rather than curriculum to help students achieve better score. Study of the washback of English as second language tests from a university in Sri Lanka discovered that undergraduate students and ESL instructors preferred to use test oriented teaching and learning materials, such as past exams, rather than reading “English Skills for New Entrants”, a free book published by the University Grant Commission (Iyer, 2015). Lodhi et al. (2018) had similar findings that English language teachers in secondary schools selected teaching materials that might help students succeed in final English language exam, such as previous tests and supplementary materials with questions in the same format as those in the final exam. English Language teachers in a Malaysian university believed that they were able to incorporate authentic materials and real-life activities into classes because of the new English Language Assessment System (ELAS) (Bokiev & Abd Samad, 2021).

In summary, the impact of tests on teaching materials, known as “textbook washback”, is very much similar to washback on the curriculum that is driven by the stakes of tests, i.e., the higher the stakes of tests, the more significant washback on teaching materials.

### ***Washback on English Teaching Methods***

Teaching methods refer to the approaches or techniques adopted by English language teachers to teach the target content and achieve learning objectives. Studies have revealed various washback on how English language teachers teach. Smith (1991) gave an exemplification of approaches teachers choose to teach towards an exam through a qualitative study of the role of external testing in elementary schools in the United States. Eight categories of exam preparation were defined as follows.

1. No special preparation. Some teachers may not have to design and adopt special activities to prepare the pupils for the test.
2. Teaching test-taking skills. Students need some skills to take tests, such as working within time limits or transferring answers to a separate answer sheet,
3. Exhortation. Teachers would encourage students to get a good night’s sleep and breakfast before the test and to try their best on the test itself.
4. Teaching the content known to be covered by the test.
5. Teaching to the test. Teachers use materials that mimic the format and cover the same curricular territory as the test.
6. Stress inoculation.
7. Practicing on items of the test itself or parallel forms.

8. Cheating. Teachers may provide students with extra time, with hints or rephrasing of words, or with the correct answers.

(Smith, 1991, pp. 526-537)

Wall and Alderson (1993, p. 127) stated that the introduction of the new English school-leaving examination in Sri Lanka “had virtually no impact on the way that teachers teach” despite teacher reported that the examination influenced their methods. Cheng (1997) revealed that, after the introduction of a revised examination, changes were only found in teaching content but not in teaching method. Furthermore, no significant change was found in number of lectures, and the lessons were generally taught the same before and after the introduction of the new exam guideline.

Other studies indicated that methods adopted to teach towards exams vary among each individual teacher. Alderson and Hamp-Lyons (1996), based on empirical data from both TOEFL and non-TOEFL classes, found test influence on teachers’ methodology, but noted that “the effect is not the same in degree or kind from teacher to teacher” (p. 295). Watanabe (1996) discovered large differences existed in the way teachers use to prepare students for the same exam in Japan. Some adopted more test-driven approaches, i.e., ‘teaching to the test’, ‘textbook slave’ methods, while others tried more creative and independent approaches. Wang (2009) observed that 8th grade English language teachers in China adopted test-oriented methods, such as reviewing previous tests with students, designing and requiring students to complete assignments that follow exactly the same formats in the English language test for high school entrance examinations. The post-observation interviews revealed that these test-oriented teaching methods negatively impacted English language teachers’ teaching interest and their students’ learning interest. Iyer (2015) also found that English language teachers had to intensively give practice tests to undergraduates who must pass the required English language exam, even though they understood this exam would not comprehensively assess the overall English language proficiency of students.

Lodhi et al. (2018) discovered majority ELL teachers in secondary schools in Pakistan only selected teaching methods that would help their students succeed in English language tests. However, English language teachers in a Malaysian University had overall positive views of intended washback from an English Language Assessment System (ELAS) on teaching methods and acknowledged they had more freedom to use various teaching techniques to make classes “creative” and to address their students’ needs (Bokiev & Abd Samad, 2021).

Overall, unlike washback on curriculum and teaching materials, washback on teaching methods varies in different contexts with individual teacher. Further empirical studies are needed to explore the diversity of washback on teaching methods.

### ***Washback on Student Learning***

Student learning is one of the key questions that teachers and educators have regarding washback. Does exam washback affect student learning? If so, how does it affect student learning? The articles reviewed provide some much needed empirical evidence about whether students have learned more or better because they have studied for a particular test.

Hughes (1988) investigated students’ performance in the Michigan Test and teachers’ perceptions of the gains from the first cohort of students to pass a new test, and noted that students’ performance improved after the new exam was introduced in a Turkish university. The findings indicated that factors like the nature of the test, criterion references, and student needs, contributed to the washback effect. After analyzing scripts and scores of NUE (New Use of English) test, Lam (1993) concluded that it brought positive washback to students learning given that the test covers a wider range of skills, e.g., the Practical Skills for Work & Study, which emphasizes students’ abilities to use English language in practice.

However, Alderson and Hamp-Lyons (1996) reported that “Powers . . . found only dubious evidence for the claims made by coaching companies and test preparation materials publishers that either courses or published materials have any significant effect on students’ SAT scores.” (p. 294) Similarly, Cheng (1997), based on the study of Hong Kong Certificate of Education Examination (HKCEE), drew a conclusion that “The washback effect of this exam seems to be limited in the sense that it does not appear to have a fundamental impact on students’ learning. For example, students’ perceptions of their motivation to learn English and their learning strategies remain largely unchanged.” (p. 297). Andrews et al. (2002) explored the effects of the introduction of a new oral component into a public exam in Hong Kong. Simulated oral tests with three groups of candidates were used to measure students’ oral performance. The results show that only small performance improvement between the first and the third group. They concluded that

“The sort of washback that is most apparent seems to represent a very superficial level of learning outcome: familiarization with the exam format, and the rote learning of exam specific strategies and formulaic phrases . . . the inappropriate use of these phrases by a number of students seems indicative of memorization rather than meaningful internalization. In these

instances, the students appear to have learnt which language features to use, but not when and how to use them appropriately.” (pp. 220-221)

Reynolds (2010) investigated ELL learners’ perspective on washback of TOEFL and found that many factors contributed to their perceptions of whether TOEFL preparation brought positive and negative influences to students’ English language learning, such as different English language proficiency, pressure due to various target scores, previous experience with the exam, and interaction with their instructors, etc. It was recommended that positive washback from TOEFL can be generated by more effective and frequent interaction between instructors and their ELL students, grouping students appropriately based on their English language proficiency, identifying students’ learning needs, and share opinions on TOEFL preparation, etc.

Study of the washback of English as a foreign language assessment on the undergraduates showed that learners committed more time preparing for the test to achieve higher mark instead of learning the language skills (Maniruzzaman, 2016). Since the communicative competence was not tested, the undergraduates did not consider it was important. Bokiev and Abd Samad (2021) studied the washback of an English Language Assessment System (ELAS) on English language teaching and learning in a Malaysian University. All the interviewed students expressed positive views about this new English language assessment because it comprehensively and effectively evaluated their overall English language skills. They mentioned several benefits brought by this new test, such as “making English learning more practical”, “offering diversified opportunities to demonstrate their English skills”, “using different forms of test to promote holistic language learning”, etc.

In summary, the findings from these washback studies focusing on student learning are also inconsistent. Further investigations are needed to generalize if the washback is positive or negative in various test settings with respect to student learning.

### ***Washback on Attitudes and Feelings of English Language Teachers and Learners***

Studies have shown washback impact on English language teachers and learners’ attitudes and feelings. While some noted their attitudes towards exam were consistent, say, both positive and negative, other revealed different feelings from teachers and students. For example, Li (1990) found that, though the introduction of the Matriculation English Test (MET) made teachers uncomfortable, a few years later, the survey revealed that “the overwhelming majority of the teachers had accepted these subtests along with the whole MET, admitting that the subtests were an effective measure of the candidates’ ability to use English” (p. 402). Students also held positive attitudes towards the exam and outside the classroom. There seemed to be a new enthusiasm for learning English such as more after class English learning.

However, more studies reported negative attitudes and feelings caused by language exams. For example, Shohamy et al. (1996) found negative feelings towards the Arabic exam and complaints of the test’s lack of importance. As for the high-stakes EFL exam, in spite of both teachers and students acknowledge its considerable importance, it generated “an atmosphere of high anxiety and fear of test results among teachers and students” and “teachers feel that the success or failure of their students reflects on them and they speak of pressure to cover the materials for the exam” (pp. 309-310). In another case study, although the exam made students work to achieve good scores, students still did not believe that exams are an accurate reflection of every aspect of their language learning (Cheng, 1997). Teachers not only felt pressure but guilty if failing to get students familiar with the test formats. Alderson and Hamp Lyons (1996) stated that most teachers had a negative attitude towards teaching TOEFL. They mentioned teachers’ feelings of time pressure and frustration at “being unable to make the content interesting or to ensure improved scores for their students” (Shohamy et al., 1996, p. 292).

Study of washback effects on English test for high school entrance examination found that both English language teachers and learners’ feelings and attitudes towards English teaching and learning were more negatively influenced by this test (Wang, 2009). “Teachers’ general negative attitudes toward ETS/TOEFL are produced by several mental factors, such as teaching interest, teaching creativity, teaching enthusiasm, and teaching effort” (p. 80). Students can hardly develop a life-long language learning skill in the “everything is test-oriented” atmosphere. English language teachers in a Malaysian university commented that the new English Language Assessment System (ELAS) had a positive impact on their teaching motivation as well as professional development (Bokiev & Abd Samad, 2021). Students with a variety of English language proficiency reported that ELAS, which adopted diversified forms of assessment, motivated them to learn English because what they learn can connect to real life situations. They shared feelings that ELAS also considerably improved their confidence in using English.

However, discrepancies do exist among teachers and students’ feelings and attitudes. Interviews of students in TOEFL preparation courses at three different institutions in the United revealed that the students’ views were different from their teachers’ regarding methods and materials in the exam preparation classes: “... Most teachers claimed that it was students who drove the methodology, who insisted on practice tests and on work on TOEFL-like items. However ... in our discussions with students we did not find these claims born out”. (Alderson & Hamp-Lyons, 1996, p. 286)

In summary, although washback on feelings and attitudes of teachers and students, either positive or negative, seems less complex than its impact on other aspects, it deserves further investigation of whether and how these attitudes promote or demote the effectiveness of teaching and learning.

### Conclusion and Implications

This study presents a broad review of representative empirical studies of washback in English language testing in the past thirty years and summarizes its diverse impact on ELL teaching and learning from six aspects: teaching contents, teaching materials, teaching methods, student learning, teachers' feelings and attitudes, and students' feelings and attitudes.

First of all, studies on language testing washback have become a major area within educational research, and English language teaching and learning in particular. Secondly, washback exists in various areas and its impact appears in different forms and degrees, which contribute the complexity of this educational phenomenon. The analysis of washback in English language tests "demands careful context-based interpretation" (Kuang, 2020, p. 15). Thirdly, different methodologies have been adopted, though not simultaneously, to study washback, such as classroom observation, individual and group interviews as well as the analysis of teaching materials, to increase the validity of the studies of washback. Finally, researchers who investigate washback effects not only have described what washback looks like, they have also attempted to explain why it appears as it does. Many suggested that issues, such as teacher and student factors, stakes of tests, and the contexts, should be taken into consideration in washback studies.

Due to the inevitable nature of washback in a test, researchers could explore the mechanisms in which what factors at both micro and macro levels could produce and mediate intended washback on English language classroom teaching while minimizing negative washback caused by the test. Such scholarly works on washback of English language tests and its impact on classroom teaching will be beneficial for English language developers, educational policy makers, English language teachers, as well as English language learners. When preparing for a well-developed English language test that comprehensively assesses ELL learners' overall language knowledge and competences, teachers are encouraged to embed test content into their daily classroom teaching in creative ways. This alignment between the test goals and learning objectives will motivate English language learners to fully focus on learning the language for practical use instead of preparing for a test, and it will help to achieve long term goals set by language teaching policy makers and test developers.

### Works Citation

- Adnan, U., & Mahmood, A. M. (2014). Impact of public examination on teaching of English: A washback perspective. *Journal of Education and Practice*, 5(2), 132-139.
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129.
- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13(3), 280-297.
- Andrews, S., Fullilove, J., & Wong, Y. (2002). Targeting washback - A case study. *System*, 30(2), 207-223.
- Azadi, G., & Gholami, R. (2013). Feedback on washback of EFL tests on ELT in L2 classroom. *Theory & Practice in Language Studies*, 3(8), 1335-1341.
- Baker, D. (1989). *Language testing: A critical Survey and Practical Guide*. London: Edward Arnold.
- Bailey, K. M. (1996). Working for washback: A review of washback concept in language testing. *Language Testing*, 13(3), 257-279.
- Bokiev, U., & Abd Samad, A. (2021). Washback of an English language assessment system in a Malaysian University Foundation Programme. *The Qualitative Report*, 26(2), 555-587.
- Carroll, J. & Hall, P. (1985). *Make Your Own Language Tests: A Practical Guide to Writing Language Performance Tests*. Oxford: Pergamon Press.
- Cheng, L. (1997). How does Washback influence teaching? Implications for Hong Kong, *Language and Education*, 11(1), 38-54.
- Hamp-Lyons, L. (1997). Washback, impact and validity: Ethical concerns. *Language Testing*, 14(3), 295-303.

- Hughes, A. (1988). Introducing a needs-based test of English language proficiency into an English medium University in Turkey. In A. Hughes (Ed.), *Testing English for University Study (ELT Documents #127)* (pp. 134-146). London: Modern English Publications in association with the British Council.
- Hughes, A. (1989). *Testing English for Language Teachers*. Cambridge: Cambridge University Press.
- Hughes, A. (1994). *Backwash and TOEFL 2000*. Commissioned by Educational Testing Service (ETS). University of Reading.
- Iyer, M. S. (2015). Impact of washback in English as second language classrooms – An investigation in the University of Jaffna – Sri Lanka. *ELT Voices-International Journal for Teachers of English*, 5(3), 88-91.
- Kuang, Q. (2020). A review of the washback of English language tests on classroom teaching. *English Language Teaching*, 13(9), 10-17.
- Lam, H. P. (1993). *Washback-Can it be Quantified? A Study on the Impact of English Examinations in Hong Kong.*, Unpublished MA dissertation. University of Leeds, UK.
- Lam, H. P. (1994). *Methodology Washback - an Insider's View*. In D. Nunan, R. Berry, & V. Berry (Ed.), *Bringing About Change in Language Education: Proceedings of the International Language in Education Conference 1994* (pp. 83-102). Hong Kong: University of Hong Kong.
- Li, S. (1997). *Language Testing Technology and Arts*. Hunan: Hunan Education Press.
- Lodhi, M. A., Robab, I., Mukhtar, S., Farman, H., & Farrukh, S. (2018). Impact of washback on ESL students' performance at secondary level. *International Journal of English Linguistics*, 8(6), 227-239.
- Madaus, G. F. (1988). The influence of testing on the curriculum. In L.N. Tanner (eds.), *Critical issues in curriculum: Eighty-seventh yearbook of the National Society for the Study of Education (Part 1)*, Chicago: University of Chicago Press.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241-256.
- Maniruzzaman, M. (2016). *EFL resting washback: Assessment of learning or assessment for learning?* Center for Pedagogy (CP), Independent University, Bangladesh (IUB), 350-369.
- Morrow, K. (1979). Communicative language testing: Revolution or evolution? In Brumfit, C. & Johnson, K. (Eds.), *The Communicative Approach to Language Teaching*. (pp. 143-157). Oxford University Press.
- Pearson, I. (1988). Tests as levers of change. In D. Chamberlain & R. J. Baumgardner (eds.), *ESP in the Classroom: Practice and Evaluation*. *ELT Documents* (Vol. 128, pp. 98-107). London: Modern English Publications.
- Prodromou, L. (1995). The backwash effect: from testing to teaching. *English Language Journal*, 49(1), 13-25.
- Qi, L. (2004). Has a high-stakes test produced the intended changes? In L. Cheng, Y. Watanabe, & A. Curtis (Ed.). *Washback in Language Testing: research contents and methods* (pp. 171-190). Lawrence Erlbaum Associates, Inc.
- Shahzad, S. (2006). Aiming for positive washback: A case study of international teaching assistants. *Language Testing*, 23(1), 1-34.
- Shohamy, E. (1993). *The Power of test: The Impact of Language Testing on Teaching and Learning*. National Foreign Language Center Occasional Papers. Washington, DC: National Foreign Language Center.
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13(3), 298-317.
- Shu, D. (2004). *Foreign Language Teaching and Reformation: Problems and Solutions*. Shanghai: Shanghai Foreign Language Education Press.
- Swain, M. (1985). Large-scale communicative testing. In Y. P. Lee, A.C.C.Y. Fok, R. Lord, & G. Low (Eds.). *New directions in language testing* (pp. 35-46), Oxford: Pergamon Press.

- Tsagari, D. (2007). Review of Washback in Language Testing: How has been Done? What More Needs Doing? Lancaster University, UK: Report (Eric Document Reproduction Service No. ED 497709)
- Wang, Y. W. (1997). An Investigation of Textbook Materials Designed to Prepare Students for the IELTS Test: A Study of Washback. Unpublished M.A. dissertation, Department of Linguistics and Modern English Language, Lancaster University, Lancaster, UK.
- Wang, H. X. (2008). Reflection on English exam standard and test structure, *Elementary and Middle School English Teaching and Research*, 216(4), 50-56.
- Wang, L. (2009). Washback of ETS HSE on Junior English Teaching and Learning. Unpublished Master's Thesis, Shandong: Shandong University of Finance.
- Watanabe, Y. (1996). Does grammar-translation come from the Entrance Examination? Preliminary findings from classroom-based research. *Language Testing*, 13(3), 318-333.
- Watanabe, Y. (1997). The Washback Effects of the Japanese University Entrance Examinations of English-classroom-based Research. Unpublished PhD thesis, Department of Linguistics and Modern English Language, Lancaster University, Lancaster, England.
- Watanabe, Y. (2000). Washback effects of the English section of Japanese University Entrance Examinations on instruction in pre-college level EFL. *Language Testing Update*, 27, 42-47.
- Watanabe, Y. (2004). Methodology in washback studies. In L. Cheng, Y. Watanabe, & A. Curtis (Ed.). *Washback in Language Testing: Research Contents and Methods*. Lawrence Erlbaum Associates, Inc.
- Xu, Q. (2004). National College Entrance Examination (NCEE) English testing and its washback effects. Unpublished Master's Thesis. Shanghai: Shanghai Foreign Language University.
- Zhu, Y. (2006). Empirical analysis of washback effects in testing. *Journal of XiAn Engineering & Technology University*, 20(2).
- Zou, S. & Xu, Q. (2017). A washback study of the test for English majors for grade eight (TEM8) in China: From the perspective of University program administrators. *Language Assessment Quarterly*, 14(2), 140-159.

### **Biography**

Ling Wang is an Associate Professor of Education and TESOL Program Coordinator at Austin Peay State University. She holds a Ph.D. degree in Literacy Studies from Middle Tennessee State University. She also holds a M.A. Ed. degree in Reading from Austin Peay State University and a M.A. degree in Foreign Languages and Applied Linguistics from Shandong University of Finance. Her research interest is in foreign language acquisition, literacy studies, and educational multimedia. She has published multiple refereed articles in journals and conference proceedings. Dr. Wang is a member of International Literacy Association, and she currently serves on the Editorial Boards of *Journal of Technology and Teacher Education* and *Journal of Educational Multimedia and Hypermedia*.