A Review Research of Washback in Language Testing

HU Xin[a],*

[a] School of Law and Humanities, China University of Mining & Technology-Beijing, Beijing, China.
* Corresponding author.

Abstract

Washback is an important concept in language testing. It reflects the effect of the testing. Based on washback, modification and optimization can be made to face the need of teachers and learners. This article aims to make a review research on washback at home and abroad to find a new direction for future research. The research is based on proper journal articles on both core international journals and CNKI. In line with all related articles, three main aspects of washback research - theoretical models, positive and negative washback effects - are illustrated. Theoretical models contribute to empirical research and help to better understand washback. Negative washback is a common phenomenon in testing so that possible reasons causing the phenomenon will be mainly discussed. Positive washback is usually generated by making some updates. This research provides an insight into a multi-method study with the other field - test ethics - in the educational system.

Key words: Washback; Language testing; Test ethics

1. INTRODUCTION

In the field of language testing, washback refers to the influence a test has on teaching and learning. As an important part of the post-test effect, washback is a common phenomenon in the educational and applied linguistics literature. The understanding of the testing effect can date back to the 1950s and 1960s. Vernon (1956) and Wiseman (1961) pointed out test distorted the curriculum so that teachers didn't focus on the explanation of language skills but students' test-taking skills. Washback wasn't treated as a crucial research field until the 1980s. Hughes (1989) stated washback was a key concern for teachers. Before the 1990s, there was almost no empirical evidence confirming washback existed in language testing (Alderson, 2011). 'Does washback exist?' was a landmark publication in its time (Alderson & Wall, 1993) and discussed whether washback existed and if it is positive or negative. A series of washback hypotheses were formulated to deconstruct the term washback, and the nature of washback was explored. Then there appeared studies on washback which used developed models to analyze different elements in it. This paper first discusses the definition of washback. Next, concentration is kept on outlining the researches carried out into washback. Washback can be positive or negative. This part will explore studies on three aspects - theoretical models of washback, positive washback and negative washback exerted through a test. Finally, a summary is made on the findings above and find a new possibility for future research on washback in language testing. With the updating of large-scale tests and deep development of tests in real life, studies on washback inevitably become more popular among researchers in the language testing field.

2. THE PHENOMENON OF WASHBACK IN LANGUAGE TESTING

In language testing, different researchers defined washback in different ways. Some researchers preferred to

use 'washback' while others used 'backwash' to describe the effects brought by tests. As for 'washback', it is a term used principally in British applied linguistics (Alderson & Wall, 1993). Pearson (1988) suggested washback refers to the impact, which made by tests, on the behaviors, motivation, and attitudes of parents, learners, and teachers. This influence occurred at the end of a course, which is in a backward direction. Hughes (1989) stated the most concise and explicit definition that washback is the influence of testing acting on teaching and learning. It was supported by later researchers (Alderson & Wall, 1993; Bailey, 1996; Cheng & Curtis, 2004). Washback can be also treated as a bond between teaching, learning, and testing (Shohamy et al., 1996, Hamp-Lyons, 1997). Messick (1996) stressed, to some extent, washback can affect teachers to do something they wouldn't have done to stimulate or suppress language learning. Moreover, washback indicates the change of curriculum function and direction, which is intended or accidental, influence teaching and learning through a reform of the test (Cheng, 2005). In terms of 'backwash', It is used in the general education circle (Alderson & Wall, 1993). Backwash depicts the control had by testing over the curriculum, learning strategies, and teaching methods (Biggs, 1995). Meanwhile, backwash can be found in dictionaries, but not washback. These two expressions can be used interchangeably (Hughes, 1993).

Except for 'backwash' and 'washback', researchers adopted some other terms to present similar phenomena in the educational field. 'Test impact' indicates the tests can have a further impact on a wider framework - educational field - rather than only in a language class. Bachman & Palmer (1996) used 'test impact' to state the influence that the tests exert on a micro level and macro level. On a micro level, tests mainly affected individuals, such as students and teachers. On a macro level, educational systems or society might be influenced by tests. Wall (1997) asserted 'test impact' as any influence tests may have on individuals, practices or policies in the classroom, the school, the educational system or society. The effect of the test on teaching and learning was rooted in 'measurement-driven instruction' (Cheng & Curtis, 2004). MDI - 'measurement-driven instruction' - stresses high-stakes tests could promote valid teaching and learning as long as they were designed and carried out properly (Popham, 1987). 'Curriculum alignment' referred to the content and format of the curriculum was encouraged to match with that of the test (Shepard, 1990). The closer they connect, the greater the potential improvement on the test. 'Systemic validity' is defined as the effect of teaching reform made by bringing in revised tests or new tests to the education system to promote teaching and learning (Fredericksen & Collins, 1989). Besides, 'Consequential validity' implied the positive or negative social consequences of testing served as part of a broader

and unified concept of test validity (Messick, 1996). 'Washback validity' is another term that researchers use to describe the effect brought by tests. It indicated the quality of the relationship between a test and related teaching; and it was presumably like that a good washback of the test meant the test is valid, and vice versa (Morrow, 1986).

To sum up, even though different researchers have different understandings of washback, there are still some commonalities. Washback mainly discusses the effect the tests have on teaching and learning within the classrooms or educational field. The effect can be positive or negative, so there are positive washback and negative washback. Moreover, washback acts on both individuals and society.

3. SEARCH METHOD

Washback is an important concept in language testing. To get some better literature abroad to review, I choose the papers on four high-quality journals in this field. I use the Duxiu Search Database (www.duxiu.com) to search English journal articles on washback. I adopt advanced search in 'journal' module and set 'washback' and 'language testing' for all fields, and then limit 'discipline' as language and 'type of journal' as 'Language Testing' or 'Language Assessment Quarterly' or 'System' or 'Language Teaching Research'. Finally, I get 43 related English articles.

Besides, I use CNKI to search for Chinese articles on this theme. In advanced search of the 'journal' module, I limit 'theme' as 'washback' and 'language testing', and limit 'type of journal' as 'Core journal' and 'CSSCI'. Then I get 45 related Chinese articles on washback in language testing.

4. MAIN RESEARCH ANALYSIS

After sorting out all the articles I get above, I divide them into three parts to give a further explanation - theoretical models of washback, positive washback exerted through tests, and negative washback exerted through tests. In terms of the research paradigm, there are both empirical studies and non-empirical studies on washback. Empirical studies are mainly used to explore the washback direction, which refers to the positive or negative effect on target objects. Non-empirical studies usually present important concepts of washback and reviews of previous findings to provide an insight into a better understanding of washback.

4.1 The Theoretical Models of Washback

Since the 1990s, studies on washback became more and more popular. In addition to discuss the definition of washback, researchers began to explore theoretical models of washback so that they can process some empirical studies to confirm the washback mechanism and the nature of washback in real situations. Alderson and Wall

(1993) proposed 15 washback hypotheses after reading literature on language testing and talking to teachers about their teaching and testing. These hypotheses were strongly supported by many later researchers and popularly used for a long time then. They argued that tests affect teaching contents, teaching rate and sequence, teaching quality and quantity, teaching and learning attitudes and methods. Besides, they stated the relationship between test consequences and washback. These hypotheses are mainly considered from two aspects - teaching (teachers) and learning (students), which lay the foundation for future empirical studies. Alderson and Wall (1993) suggested later researchers give thought to research findings in at least two areas - motivation and performance, innovation and change in the educational field. They focused principally on micro-aspects of the teaching and learning process, which might be influenced by tests. Furthermore, they confirmed a strong relationship existed between test significance and the extent of washback through these hypotheses.

In addition, Hughes (1993) stated that a more precise explanation is needed to illustrate what constituted washback. He came up with trichotomy to reveal how washback worked in various contexts. The trichotomy refers to participants, process and product in teaching and learning, and all three may be influenced by the nature of tests. Among them, 'participants' include students, classroom teachers, administrators, materials developers and publishers; 'process' refers to any action taken by participants which is conducive to learning process, such as materials development, changes in teaching methodology, and test-taking strategies; and 'product' means learning contents, such as skills and facts, and the quality of learning, like fluency. A test may first affect participants' perceptions and attitudes which can then influence what participants undertake during the process. This might finally affect learning outcomes. Based on trichotomy and 15 washback hypotheses, Bailey (1996) built a new model of washback and presented the complexity of the washback mechanism. She added 'researchers' into 'participants', and stressed how the products might influence other products. She defined the direct effect of test-derived information exerted on the test-takers as 'washback to the learners'; and the results of test-derived information provided to other participants as 'washback to the program'. In Bailey's opinion, a test directly affects the participants within various processes, which results in products specific to each category of participants. She pointed out the participants might react to the test, which is absent in Hughes' model. Both Hughes' trichotomy and Bailey's new model of washback focus on the effect of a test in every category of the educational system.

Later, Green (2006) built a new model that is more comprehensive to explain washback effects. Compared to previous models, Green's model has a stronger explanatory power, for washback direction is added. There are three main parts in this model - washback direction, washback variability, and washback intensity. In the washback direction, it discusses the overlap between test design characteristics and constructs validity. The test designs are most closely related to the washback direction, which determines whether a positive or negative effect on teaching and learning. Washback variability involves the characteristics and values of participants. The characteristics and values include an understanding of test requirements, acceptance of test requirements, resources to meet test requirements. These may lead to different effects on different stakeholders in the same test. The perception of test importance and difficulty provides a premise to washback intensity. The more important the tests are, the more intense the washback is. However, the test difficulty doesn't work in this way. Only the challengeable and moderate difficulty can trigger the most intense washback.

Except for Green's model, there is another new model of washback later. Shin's (2007) model considered more factors and stressed the effect on students brought through the interaction between multiple factors and the test. Shin proposed this washback model by using three categories - extrinsic, intrinsic, and test factors. The model presents that these three factors determine the washback of testing on the learning and personal psychology of students. In Shin's research, extrinsic factors may have an impact on intrinsic factors, and the results of the test can affect intrinsic factors. Different from washback hypotheses and Bailey's model, Shin's model stresses on the washback of tests on learning and describes the complexity of it in detail. It also complements washback hypotheses and Bailey's model by identifying the influence of society, family, and personal affairs on students' learning. Later, Shin (2010) also established a similar model of washback to illustrate the factors acting on the influence of testing on the policies of departments or schools.

For the previous models, they are built according to some categories in the language field. However, Xie and Andrews (2013) proposed a model of washback with structural equation modeling which is a concept in another field. Structural equation modeling is a comprehensive statistical approach to deal with data to analyze the relationship between variables. Xie and Andrews adopted expectancy-value motivation theory to illustrate the paths of impacts from the perceptions of test uses and design to test preparation, which can serve as a special case of washback on learning. Expectancy-value motivation theory, which is a concept in learning psychology, is used as a method to explain the washback mechanism on learning during test preparation. The model of washback on learning based on this theory will be verified statistically through structural equation modeling.

The models of washback mentioned above have been further improved in turn, that is, the latter one is more comprehensive and proper than the former one. Though the update and improvement can be seen in this aspect, there is still a great space for researchers to propose new models in line with practical situations to realize different research aims. For example, in future research, the researchers can establish a model of washback combined with the concepts in SLA to explore the effect a test brought on learning strategies in SLA.

4.2 Positive Washback Exerted Through Tests

Positive washback is not a natural effect to get through tests, so some actions need to be taken to ensure the positive influence on target objects. It's a big point to consider how to achieve positive washback through updating tests. As for the studies in this part, they can be illustrated in three aspects - positive washback on classroom assessment, positive washback on standardized testing and positive washback on student learning. The first two aspects mainly focus on both teaching and learning process, whereas the third one is only perceived from students' angle, and discusses their reactions to the testing changes and the effects generated in this process.

In general, there are two kinds of tests - high-stakes test and low-stakes test - under the impact of the test. The high-stakes test refers to the test that will make a big difference in stakeholders, such as admission, graduation, employment, and promotion. That is, it will influence the stakeholders' future. The standardized test is a typical example. However, a low-stakes test is a kind of test that will not exert a decisive effect on stakeholders. Classroom assessment is a good example.

To explore the washback on classroom assessment, researchers did some research on the effects of specific tests on classroom contexts. Muñoz and Álvarez (2010) made an empirical study to investigate the influence of an oral assessment system exerted on teaching and learning in the EFL classroom. They adopted quantitative and qualitative research methods and set a comparison group and an experimental group to make a comparison among 110 college students. The result presents positive washback on some parts of teaching and learning in the classroom. Several factors cause this result. Teachers accept constant training on assessment practices, while students are well-informed of assessment procedures, scoring scales, specific learning objectives, and well-constructed assessment tasks. Then teachers and students build a connection between educational goals and assessment, which will generate positive washback on teaching and learning in the EFL classroom. Self-assessment is another way to form positive washback to the students, for they can control the assessment by themselves. Therefore, the positive washback can support the oral assessment system to go further in a classroom assessment context. Another example to illustrate this

aspect is the Story Retelling Speaking Test (SRST) developed for classroom speaking assessment (Hirai & Koizumi, 2009). Story retelling is a speaking activity widely used by teachers in speaking class, but few test designers attempt to establish a test like SRST. Due to the features of Story Retelling activity, students are familiar with its pattern and requirements. It's easier for students to prepare the test and attend the test. Meanwhile, it can build up positive washback to the speaking assessment and classroom practices. Through the questionnaire, positive washback is also formed on students' perception of the test usefulness and the reasonableness of the test procedures. In addition, another research presented the washback effect from the teachers' point of view. The role teachers play and the decisions they make can influence the types and intensity of washback generated from tests in the classroom context (Spratt, 2005).

Except for positive washback on classroom assessment, positive washback on standardized testing is another important aspect of studies. Standardized tests always draw great attention from learners at home and abroad. To get the positive washback on teaching and learning of English, many standardized tests are optimized by test designers and relevant education departments, such as GEPT in Taiwan and NMET in China (Wu, 2012; Cheng & Qi, 2006). CET-SET (College English Test-Spoken English Test) is a standardized English spoken test, which is used to assess the English spoken proficiency of college students in Chinese universities. More explicit information on the CET-SET construct and rating scale will contribute to the generation of positive washback on teaching and learning (Zhang & Elder, 2009). CET (College English Test) is another important standardized English test for college students in Chinese universities. Zhou and Xiao (2016) took an empirical study to investigate the correlation between CET and CET-SET scores. They found that these two tests' scores had a correlation with each other, which meant CET-SET, to some extent, could reflect the real proficiency level of English. It exerted some positive washback on foreign language teaching and learning. In Chinese, NMET (National Matriculation English Test) is the most important and influential test, which will determine whether a high school student can enter a university. SHMET carried out reform on its listening and speaking test, which referred to the score of this test will be counted into the score profile of SHMET. Hou (2018) studied the washback of this reform through quantitative and qualitative methods. She stated that most teachers had a positive attitude towards the reform and it could have a positive effect on English teaching in high school. Moreover, raters' negotiation on scoring consistency in assessment can also generate positive washback on teaching practices (Trace et al, 2017).

Washback on student learning is another key consideration in washback effect studies. CET 4 is the most influential test for undergraduate college students,

which is mandatory for them. Since 2004, it experienced a large-scale reform in two major aspects. Firstly, it adjusted component weighting and counted the component scores into the score profile. Secondly, it added more open-response questions and reduced multiple-choice items. Xie (2015) investigated students' perceptions of these two changes and the impact of the changes had on time management, test preparation and test performance of students. The study found students' perceptions had bits of significant effects on listening practice, but not on reading and writing. Their favorable perceptions on CET 4 reform exerted positive washback on the learning approach. Besides, the second change enhanced test validity, which also can increase positive washback. Another researcher did similar research (Shi, 2010). She studied the washback of the new CET 4/6 listening test. The research investigated the influence of the test on learners' learning attitude, learning time, the width and depth of learning. The result showed the new test, to some extent, exerted overt positive washback on the listening learning of students. The increase in listening score weight promoted students to spend more time practicing, so their listening skills got improved.

From all the studies above, they all present that positive washback can't generate naturally. Something needs to be done or paid attention to ensure it occurs in proper position and time. However, there is little research investigating how tests affect how teachers teach in the classroom and whether that exerts a positive washback. If not, it's a good point to investigate what they can do to achieve a positive washback.

4.3 Negative Washback Exerted Through Tests

Negative washback is a common phenomenon brought by tests. In this part, two aspects will be illustrated - negative washback on classroom assessment and negative washback on standardized testing. For negative washback, the studies mostly focus on analyzing the reason lead to it.

As for negative washback on classroom assessment, it still stresses on the low-stakes test. A study investigated translation testing (from L1 to L2). It found translation would often exert negative washback on classroom practice, and it should be used carefully (Buck, 1992). It would also lead to activities that are not beneficial to SLL. Besides, another study concentrated on the negative washback of a speaking test as a final test in universities and analyzed its reasons (Liao, 2010). Due to the indifference on speaking test at the end of the term, teachers and students didn't pay much attention to teaching and learning. They spent less time practicing speaking, so their speaking skills would not be improved. The reasons for the negative washback are the limitation of teachers, testing time, testing skills, reliability and validity of the test, and bad management.

Negative washback on standardized testing is also a popular theme to study. It mainly involves some decisive high-stakes tests. Qi (2005) researched NMET and figured out why the test failed to achieve expected washback. There are two main reasons - the selection function and the function to promote changes. They are inconsistent with each other, so it will lead teaching as a way only to achieve in the test, which is not the expectation of policymakers. Moreover, Choi (2008) studied the standardized EFL test in Korea. Most of the students and teachers held a negative attitude on this test, even though it's a decisive high-stakes test. The majority of students couldn't bear the pressure from this test. They also thought there was a mismatch between the test scores and their real English proficiency. That is, The EFL test had a negative washback on EFL teaching.

To sum up, the nature of a test determines that it will leave some pressure on the participants. The test scores are often used for different purposes. Low-stakes test acts on checking learning outcomes in a period of time, while the high-stakes test often serves as a judge to learners' future. Teachers and learners will naturally choose to teach and learn based on the objective - achieving in the test. So, most of the studies in this part concentrate more on test negative effects on teachers and learners, and the teaching and learning process. There are few studies analyzing the reasons for negative washback in a wider range - the educational system. That is, test ethics can be added into research, such as fairness, differential item functioning, the responsibility of language testers and policymakers. In recent two years, this kind of multi-method studies gradually appear (Macqueen et al, 2019; Zheng, 2018), but there is still a large space to explore in future research.

5. CONCLUSION

This study mainly researches the washback effect in language testing at home and abroad. Washback is an important and popular research theme in language testing. Different researchers give different definitions of the term 'washback'. Washback can be positive or negative, acting on classroom assessment and standardized testing. To exert positive washback, some useful actions will be taken on tests to achieve it. However, the studies on negative washback often focus on the analysis of its reasons. At the beginning of washback research, non-empirical research played a dominant role. With the development of its theoretical framework, some theoretical models of washback were established by researchers. They provided great support to latter empirical research. However, there are still some limitations to washback research. The researcher before paid much attention to teaching and learning in a narrow scope, such as the classroom. The studies in the future can explore washback in a wider scope, such as the educational system, and also combine with other fields, like test ethics. Through washback research, the test can be optimized better and better to benefit teachers and learners in the teaching and learning process.

REFERENCES

Alderson, J. C. (2011). *A lifetime of language testing*. (Applied linguistic series). Shanghai: Shanghai Foreign Language Education Press.

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, *14*(2), 115-129.

Bachman, L. F., & Palmer, 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*(3), 257-279.

Biggs, J.B. (1995). Assumptions underlying new approaches to educational assessment: implications for Hong Kong, *Curriculum Forum*, *4*(2), 2-22.

Buck, G. (1992). Translation as a language testing procedure: does it work? *Language Testing*, *9*(2), 123-148.

Cheng, L. (2005). *Changing language teaching through language testing: A washback study (Vol.21)*. Cambridge, UK: Cambridge University Press.

Cheng, L., & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in Language Testing: Research Contexts and Methods* (pp. 3-17). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Cheng, L., & Qi, L. (2006). Description and examination of the national matriculation English test. *Language Assessment Quarterly*, *3*(1), 53-70.

Choi, I.-C. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing*, 25(1), 39-62.

Green, A. (2006). Watching for washback: Observing the influence of the international English Language Testing System Academic Writing Test in the Classroom. *Language Assessment Quarterly*, *3*(4), 333-368.

Hamp-Lyons, L. (1997). Washback, Impact and Validity: Ethical Concerns. *Language Testing*, *14*(3), 295-303.

Hirai, A., & Koizumi, R. (2009). Development of a practical speaking test with a positive impact on learning using a story retelling technique. *Language Assessment Quarterly*, *6*(2), 151-167.

Hou, Y. (2018). A study on the washback effect of the reform of SHMET listening and speaking test. *Technology Enhanced Foreign Language Education*, (05), 23-29.

Hughes, A. (1989). *Testing for Language Teachers*. Cambridge: Cambridge University Press.

Hughes, A. (1993). *Backwash and TOEFL 2000*. Unpublished manuscript, commissioned by Educational Testing Service (ETS). University of Reading.

Liao, Y. (2010). Reflection on the failure of oral english test in college English final examination. *Education and Vocation*, (02), 178-179.

Macqueen, S., Knoch, U., Wigglesworth, G., Nordlinger, R., Singer, R., McNamara, T., & Brickle, R. (2019). The impact of national standardized literacy and numeracy testing on children and teaching staff in remote Australian Indigenous communities. *Language Testing*, *36*(2), 265-287.

Messick, S. (1996). Validity and washback in language testing. *Language Testing*, *13*(3), 241-256.

Morrow, K. (1986). The evaluation of tests of communicative performance. In: M. Portal (Ed.), *Innovations in Language Testing* (pp. 1-13). London: NFER/Nelson.

Muñoz, A. P., & Álvarez, M. E. (2010). Washback of an oral assessment system in the EFL classroom. *Language Testing*, *27*(1), 33-49.

Pearson, I. (1988). Tests as levers for change. In D. Chamberlain & R. J. Baumgardner (Eds.), *ESP in the Classroom: Practice and Evaluation. ELT Documents Volume 128* (pp. 98-107). London: Modern English Publications.

Popham, W. J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappan*, *68*(9), 679-682.

Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing*, *22*(2), 142-173.

Shepard, L. A. (1990). Inflated test score gains: Is the problem old norms or teaching the test? *Educational Measurement: Issues and Practice*, *9*(3), 15-22.

Shi, X. (2010). A longitudinal backwash study of the new CET 4/6 listening tests. *Foreign Language World*, (03), 80-86.

Shih, C.-M. (2007). A new washback model of students' learning. *Canadian Modern Language Review*, *64*(1), 135-161.

Shih, C.-M. (2010). The washback of the general English proficiency test on university policies: A Taiwan case study. *Language Assessment Quarterly*, *7*(3), 234-254.

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

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.

Trace, J., Janssen, G., & Meier, V. (2017). Measuring the impact of rater negotiation in writing performance assessment. *Language Testing*, *34*(1), 3-22.

Vernon, P. (1956). *The measurement of Abilities*. London: University of London Press.

Wall, D. (1997). Impact and washback in language testing. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education. Vol. 7. language testing and assessment* (pp. 291-302). Dordrecht: Kluwer Academic.

Wiseman, S. (1961). The efficiency of examinations. In Wiseman, S (ed). *Examinations and English Education*. Manchester: Manchester University Press.

Wu, J. R. (2012). GEPT and English language teaching and testing in Taiwan. *Language Assessment Quarterly*, *9*(1), 11-25.

Xie, Q. (2015). Do component weighting and testing method affect time management and approaches to test preparation? A study on the washback mechanism. *System*, *50*, 56-68.

Xie, Q., & Andrews, S. (2013). Do test design and uses influence test preparation? Testing a model of washback with Structural Equation Modeling. *Language Testing*, *30*(1), 49-70.

Zhang, Y., & Elder, C. (2009). Measuring the speaking proficiency of advanced EFL learners in China: The CET-SET Solution. *Language Assessment Quarterly*, *6*(4), 298-314.

Zheng, Y. (2018). Washback and test ethics based on a unified view of validity — Reconstruction of the participants relation in English Test. *Journal of North Minzu University (Philosophy and Social Science)*, (04), 156-160.

Zhou, S., & Xiao, F. (2016). An empirical study of correlation between CET and CET-SET Scores: based on washback theory. *Foreign Languages Research*, *33*(02), 52-57+112.