



The Advantages and Disadvantages of Quantitative and Qualitative Approach for Investigating Washback in English Language Testing

Venny Karolina¹✉, Muaffi Alif², Sudharni Sudharni³

Tanjungpura University, Indonesia¹

Adicita Mulia Primary School, Indonesia²

Aisyiyah Midwifery Academy Pontianak, Indonesia³

E-mail : karolina_venny@yahoo.co.uk¹, muaffialif10@gmail.com², ninisudharni@gmail.com³

Abstrak

Makalah ini bertujuan untuk menganalisa manfaat dan kendala beberapa pendekatan penelitian untuk menyelidiki washback dalam tes Bahasa Inggris. Makalah ini terdiri dari eksplorasi konsep washback dalam tes Bahasa Inggris. Washback adalah pengaruh tes terhadap pembelajaran dan pengajaran. Gagasan ini menunjukkan bahwa penelitian washback berkaitan dengan hubungan sebab akibat. Hubungan sebab akibat dapat dapat diteliti melalui pendekatan kuantitatif, kualitatif, atau metode campuran. Setiap pendekatan memiliki kelebihan dan kendala. Makalah ini menyimpulkan bahwa pendekatan kuantitatif dan kualitatif berkontribusi pada penyelidikan washback dalam tes Bahasa Inggris. Namun pendekatan tersebut juga memiliki beberapa kendala. Penulis menyimpulkan bahwa pendekatan kualitatif lebih cocok untuk menyelidiki washback dalam tes Bahasa Inggris. Pendekatan ini dapat menangkap informasi mendalam tentang bagaimana dan mengapa guru dan siswa dipengaruhi oleh tes Bahasa Inggris. Oleh karena itu, pendekatan metode campuran melalui kualitatif dan kuantitatif direkomendasikan untuk mendapatkan hasil penelitian yang komprehensif dalam penelitian washaback pada tes Bahasa Inggris.

Kata Kunci: washback, tes Bahasa Inggris, penekatan kualitatif, pendekatan kualitatif, studi metode campuran.

Abstract

This paper aims to analyze the benefit and constraints of approaches for investigating washback in English language testing. This paper starts with the exploration of the concept of washback in English language testing. The notion of washback is related to the effect of testing on learning and teaching. This notion shows that washback research is associated with a causal relationship. A causal relationship might be investigated through quantitative, qualitative, or mixed-method approaches. Each approach may have strengths as well as constraints. This paper concludes that quantitative and qualitative approaches contribute to the investigation of washback in English language testing. However, each approach also has some constraints. This writer concludes that a qualitative approach may suit well to investigate washback in English language testing. This approach may capture in-depth information on how and why teachers and students are affected by language testing in their teaching and learning process. However, the quantitative approach may excel in many participants so that the research results may be generalized. Each approach may contribute to washback research in English language testing; therefore, a mixed-method approach: qualitative and quantitative, is recommended to gain comprehensive findings for investigating washback in English language testing.

Keywords: washback, English language testing, quantitative approach, qualitative approach, mixed-method study.

INTRODUCTION

Washback is the effect of testing toward teaching and learning (Hung, 2012). This concept comes from the idea that testing or examination can drive the way teachers teach and students learn (Taqizadeh and Birjandi, 2015). It indicates that the power of testing can determine what happened in the classroom (Hung and Huang, 2019). The term washback refers to the backward direction; testing at the end of the program is influenced by teaching and learning. However, it might have reversed meaning; testing leads to teaching and learning practices in the classroom (Phelps, 2016). This confirms that testing is influential; despite its position at the end of the course, testing might affect the direction of teaching and learning practices, including in English language learning classroom.

Some past studies have investigated the phenomenon of washback and proven the existence of washback in language testing. The finding was that the phenomenon of washback is complex as it relates to many aspects: testing, teachers, students, attitudes, behaviors, curriculum, policy, parents, and society demands (Ali and Hamid, 2020). Due to its complexity, the investigation of washback was done in many ways, ranging from observing teaching and learning practices in the classroom, interviewing the stakeholders like teachers, students, and parents to conducting the survey. Therefore, there are variations in the investigation of the phenomenon of washback. Some washback researchers use quantitative, qualitative, or combined approaches. This paper will evaluate how both approaches, quantitative and qualitative, examine the washback phenomenon. This paper will also investigate which approach that suits best and gives more advantages in the washback investigation. This writer analyzed two past studies using a quantitative and mixed-method study (quantitative and qualitative approach). The analysis was arranged based on the following organizations: an exploration of a washback concept, description of past washback studies, the paradigm of research, and the analysis of two studies of washback.

This writer initially explored the conceptual framework of washback to give an insight into washback. A conceptual framework represents a concept that shows the relationship of elements concerning the concept. This representation can be conceptualized in a diagram, graph, or narration (Punch and Oancea, 2014). This framework is essential to help the researcher determine the research area, bring clarity to research, and construct the research questions (Punch and Oancea, 2014). Considering its importance, the writer of this paper explored the conceptual framework of washback in language testing in narrative form based on past studies.

Researchers claimed that testing has an influence (Hung and Huang, 2019). It influences a broad scope of systems: society, educational system, and individuals or stakeholders within the systems, such as teachers, students, and parents; this refers to the impact of testing (Polesel, Dulfer and Turnbull, 2012). Meanwhile, the influence of testing on teaching and learning is called washback (Punch and Oancea, 2014). As the scope of washback is narrower than impact, washback is considered as one of the aspects of testing impact (Tsagari and Cheng, 2017).

(Baksh, Sallehudin and Hamin, 2019) used their own terms micro-level and macro-level effect for washback and impact respectively. Albeit the discrepancy of the terms used, Tsagari, and Cheng (2017) and Baksh et al. (2019) have similar concepts that washback refers to more specific effects: teaching and learning, while impact covers broader effect: any related parties influenced by testing. In short, the term washback is referred to the influence of testing on the way teachers teach, and students learn.

Based on the definition, washback is closely related to various aspects: test, teaching, and learning; this means that the phenomenon of washback is complex as it relates to several components (Green, 2013). Therefore, due to its complexity, the construct of washback is debated among researchers. Asma, Sabeen, and Isabel (2014) claimed that washback comprised two broad categories: the strength of testing and the direction (positive or negative) of washback. He further said that the strength or the intensity of the test would influence

classroom activities; the more important the testing is, the more effort teachers and students make to meet the testing demand. Consequently, teachers and students employ specific strategies in the teaching and learning process to succeed in testing, but the strategies are not always conducive for language learning. For example, teaching the English test by practicing test worksheets instead of teaching the English language will lead to negative washback. Meanwhile, positive washback contributes to promoting learning, for example, English speaking performance test that stimulates the English communication practices in the classroom.

By contrast, Sultana (2018) stated that the washback construct is based on the partnership between the testing and teaching; good quality of testing will result in good teaching practices, indicating that positive washback correlates to testing validity. In detail, the researcher posited that a language test that does not directly measure the skill intended will cause harmful washback to language learning practices. For example, suppose the English language test applies multiple-choice questions to assess writing skills in English. There is a tendency that classroom activities are about practicing to solve the multiple-choice questions instead of practicing writing. Nevertheless, if the test assesses directly what it is supposed to measure, positive washback will occur. For example, a performance test to assess speaking skills in English will lead to communication practices in English in the class.

Beikmahdavi (2016) supported Sultana's idea (2018) that washback is related to the validity of testing. The researcher elaborated the concept of the validity of testing into two aspects: authenticity and directness (Beikmahdavi, 2016). Authentic language testing that engages tasks closely related to real-life use would affect teachers to create real-life practices of language learning in the classroom, for example, assessing English speaking skills through simulation of debate. Furthermore, a direct assessment that assesses language skills directly influences teachers to create similar language teaching practices in the classroom by giving chances for students to perform their language skills. An example of direct assessment is that English speaking proficiency is assessed through the interview, debate, or roleplay (Mustika and Lestari, 2020). Meanwhile, an example of non-direct assessment is that English speaking skill is assessed through the ability of students to match the correct responses of conversation in the written test. Language assessment that includes authenticity and directness would contribute to positive washback, promoting the performances of language skills in the class. However, if the washback is negative, it is not certainly caused by an invalid test (Hung, 2012). A poor test might have a good effect if the teachers prepare the lesson thoroughly or students find the test necessary, so they learn seriously. This means that there might be other factors that influence washback as it is a complex phenomenon that is not directly caused by the validity of the test.

In conclusion, washback is the influence of testing on individuals concerning teaching and learning activities. The effect has two directions: positive or negative. Positive influence can be achieved if language testing is closely related to the skill of language assessed. However, suppose a language test assesses the indirect mastery of language. In that case, the washback can be negative as the classroom activities might be tailored to the teaching to testing, how the students can answer the questions, not how to use the language. Meanwhile, the intensity of influence depends on the consequences of testing to the individuals in the testing context; high stakes testing, for example, requires a specific score as passing grades will force students to work hard in their learning.

The notion of washback is closely related to influence or effect indicating causal relationship; meaning, washback investigation can be conducted in either quantitative or qualitative approach (Cheng and Curtis, 2016). However, this writer suggests that washback is examined in in-depth research. It is not only because of the complexity of the phenomenon of washback but also the complexity of classroom activities. Many factors can influence the teaching and learning process in the classroom; so, classroom activities might not be entirely affected by the testing. For example, suppose the language testing relates to writing skills, while the teacher uses writing activities less than other skills. In that case, it is not necessarily indicating that the testing does not influence teaching and learning. Other factors might influence the teacher, such as lack of training or lack

of writing competence. Therefore, this writer suggests that washback can be deeply understood through an approach that can answer how and why washback occurs, not merely whether washback exists or not.

Paradigm is a way for researchers to see the world's reality (Mertens, 2014). Paradigm is also defined as a set of beliefs that guides researchers to take action in their research (Kivunja and Kuyini, 2017). Meaning, paradigm influences how a researcher studies and interprets the knowledge (Kivunja and Kuyini, 2017). In detail, the paradigm consists of three fundamental questions related to research: what the reality is (ontology), the relationship between the reality and researcher (epistemology), and how or what method to find/study the reality (methodology) (Punch and Oancea, 2014). Lincoln, Lynham, and Guba (2011) classified the paradigm into four labels. They are positivism/post-positivism, constructivist, transformative and pragmatic. However, based on the philosophy of the nature of knowledge and how the knowledge is acquired (epistemology), these paradigms are seen under two approaches: quantitative and qualitative (Muijs, 2013). Quantitative and qualitative methods are associated with positivism/post-positivism and constructivism (Punch and Oancea, 2014).

Positivism is based on the belief that the world is natural; the social world is also natural; cause and effect can be explained; and knowledge is valid, objective, and accurate. In other words, positivists see the world as the absolute truth that can be generalized, and the truth is independent of the assumption of the researcher. Meanwhile, post-positivism also valued objectivity and generalizability, but multiple realities should be captured to understand the truth (Abu-alhaija, 2019) as the world is ambiguous and complex. Post-positivism acknowledged the validity of qualitative, but in terms of data collection and data analysis, post-positivism used a quantitative approach, which is similar to positivism (Punch and Oancea, 2014). In contrast to positivism/post-positivism, which values objectivity, constructivists see that people construct reality. A researcher can understand reality and experience from people's opinions (Jung, 2019). Constructivists relate to subjectivity as they can gather data from subjects' opinions.

Because these two paradigms have two fundamental differences in how to see reality, these approaches have different methodologies or ways to study reality. The methodologies used are associated with how the researchers know the reality. Positivists know reality as an absolute truth that should be verified. Therefore, the quantitative approach deals with theory verification (Punch and Oancea, 2014). The methods used to verify theory are measurements, describing phenomena description, experiments, and surveys. Meanwhile, constructivists understand the reality in the world through people's opinions, so qualitative approaches relate to theory generation (Punch and Oancea, 2014) and the methods used are observation, interview, experience, and interpretation.

After exploring the paradigm, the writer will evaluate the paradigm of each study of this paper. The researcher of paper 1 investigated the effect of the new exam on students' learning activities by using the method of science. The researcher measured students' perception of their learning activities and investigated the differences between them with old and new exams. In this way, the researcher studied the effect of the new exam by comparing the discrepancy of students' behavior under both circumstances. The researcher viewed the reality of washback as a real phenomenon that was assessed objectively. The researcher did not make her own interpretation in understanding students' behavior in the classroom. Instead, students themselves show their tendency by responding to the questionnaire. This researcher then is considered positivist.

Paper 2 viewed the phenomenon of washback in the classroom can be investigated through direct observation and interview rather than through a questionnaire. The researchers studied the existence of washback in the school's natural setting, indicating that the researchers view washback as the reality constructed by individuals. The reality can be interpreted by observers who see direct the activities in the classroom and by interviewers who listen to teachers and students about the effect of testing in their teaching and studying. This means that the researchers valued the constructivism position. They believe that the

concept of washback can be understood holistically in the actual context of the school (Arthur, 2012). However, although the researchers of study 2 embraced constructivism, they quantitatively analyzed the observation checklist. This means that the researchers also value positivism. Punch and Oancea (2014) suggested that combining quantitative and qualitative approaches is in line with a pragmatic paradigm that rejects assessing reality by a single method. Pragmatists apply mixed method research to solve a problem or to understand the truth of the phenomenon. Thus, the researchers of paper 2 might be considered pragmatists.

Both studies have the same investigation: washback in language testing, but they perceive differently about the concept of washback. The researcher of paper 1 considered washback as something that causes an effect (Creswell, 2013) that can be investigated through a value-free method to naturally get the causal effect explanation without the researcher's intervention (Mertens, 2014). However, this investigation of washback was then conducted separately from the classroom, where the washback existed. Therefore, the researcher can only get the changes of behavior from the result of quantitation. Anti-positivists criticized this by perceiving that life relates to "inner experience, individuality, freedom" (Cohen, Manion, and Morrison, 2013, p. 14) rather than a measurable picture. By contrast, the researchers of paper 2 investigated the washback from an actual situation where washback existed. Although the researchers analyzed the observation quantitatively, they observed the classroom activities directly. They also interview teachers and students to get information on why they have certain behavior in the classroom in the response to the English language testing. In this way, the researcher can depict the behavior of students revealed in the classroom and get the reasons why they behave in certain ways in the classroom. (Cohen et al., 2013) said that this approach could let the researcher understand the whole of a phenomenon in real life. However, objectivity will not be captured (Austin and Sutton, 2014) as the researcher's interpretation of phenomena can be interpreted differently by other researchers.

METHOD

To explore whether the quantitative or qualitative approach is best applied in the washback investigation, this writer analyzed two papers of washback in language testing that used different methods. The analysis begins with the description of each study of washback in language testing.

Paper 1:

Cheng, L. (1998). Impact of a public English examination change on students' perceptions and attitudes toward their English learning. *Studies in Educational Evaluation*, 24(3), 279-301.

This is a two-year longitudinal quantitative study that investigated the effect of new English exams on classroom activities, language practice opportunities, and students' learning strategies by using surveys of students' perceptions. The study compared the surveys administered between two cohorts of secondary students in Hong Kong; one cohort experienced the old exam in 1994 (844 participants) while another cohort had a new English exam in 1995 (433 participants).

Paper 2:

Wall, D. & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*, 10(1), 41-69.

This longitudinal mixed-method (qualitative and quantitative) study examined the effect of new English exams in 18 secondary schools in Sri Lanka using a two-year observation from 1990 to 1991. This study was arranged into two main parts: baseline observation in 1988 and observation program 1990 – 1991 (6 rounds). The baseline observation was conducted to depict what the teaching looked like before the introduction of the exam. Meanwhile, during six rounds of the observation program, the observer examined the effect of the new exam on teachers' strategies. Then, the strategies were compared before and after the introduction of a new assessment. Besides observation, the study also employed analysis of tests and group as well as individual interviews. To analyze whether washback in language testing is well suited to quantitative or qualitative

approaches, this writer evaluated paper 1 and paper 2 based on whether the findings and conclusion are trusted or called research validity. Therefore, the analysis is based on assessing research validity adapted from (Lauer, 2013): identifying research questions, confirming whether research designs suit research questions or not, and analyzing the research method.

RESULT AND DISCUSSION

As mentioned in the method session, this paper analyzed past studies related to washback in language testing based on three categories: research questions identification, research design analysis, and research method discussion. The followings are the result and discussion of this review paper. The first analysis relates to the research questions of each study. Paper 1 investigated whether any washback occurred in implementing the new English exam Hong Kong Certificate of Education Examination in English (HKCEE) that could be observed in the teaching process of secondary schools. Paper 1 also investigated how the learning process was influenced by this new exam. Paper 2 investigated the effect of the new Sri Lanka English exam on English language teaching in secondary school. Both studies investigated the effect of language exams on language teaching and learning; therefore, both have similarities in what they investigated: causal relationship.

The second analysis is to confirm that the research questions addressed match to research design chosen. Both studies have the same investigation of causation: washback in language exam, but they have different approaches: quantitative (paper 1) and mixed-method approach: qualitative and quantitative (paper 2). There were some past debates on the best approach to examine causation. A cause-and-effect relationship might be better explored with a quantitative than a qualitative approach (Muijs, 2013) as a quantitative approach can identify what causes some events with a standard format. In this way, the approach provides a clear answer. Furthermore, some qualitative researchers are reluctant to assess causal investigation as they claim that life is very complex and changing rapidly. Therefore, it is complicated to precisely claim causation as many factors can cause effects; this adds that causation is closely related to quantitative connotation (Neuman, 2011).

However, Lichterman and Reed (2015) also claimed that the theories of causation are common in ethnography. The researchers investigated factors that make changes in a group of people. The causal investigation in ethnography is even more comprehensive as it examines variable causes changes and describes what makes the changes happen (Small, 2013). Similarly, Sykes et al. (2019) stated that the qualitative approach is well suited to a causal relationship. The researcher can directly look at the causation and process of why and how it happened. This opinion brings the logical rationale that the qualitative approach is powerful to examine causal relationships as the researcher has a closer look than the quantitative approach, which analyses cause-effect relationships using calculation.

Regardless of debates, each approach has its strength or advantage in constructing causal inferences (Palinkas, 2014); each approach has its specific role in research. A qualitative approach is essential as the early stages of inquiry provide an in-depth understanding of the cause and effect that exists in the context. By contrast, the quantitative approach will be best suited to confirm the early findings and generalize findings in another context. In conclusion, either quantitative or qualitative methods might be used by paper 1 and paper 2, to answer the research questions. In other words, both approaches are applicable and valuable in causal inquiry. They even have complementary roles that support one another. The qualitative approach can be used to obtain the reality of the effect of testing in the classroom; this refers to theory generation (Punch and Oancea, 2014). The quantitative approach can be used to test whether the finding is significant and reliable; this is known as theory verification (Punch and Oancea, 2014).

In terms of research design, both studies used the longitudinal study to seek the impact before and after the new exam application. A longitudinal study is a set of studies that are conducted during a period of time (Cohen, Manion, and Morrison, 2013). In detail, longitudinal studies may involve the same sample or new

sample over time. It enables the researcher to notice the similarities, differences, or changes over a period of time (Caruana et al., 2015). This kind of design allows the researcher to establish inferences on causality that happened over time. Paper 1 did a two-year longitudinal study using a questionnaire to compare students' perception between two cohorts: old (1994) and new (1995) exams. In this way, the researcher can collect data on the gap of learning activities before and after the new exam, so the existence of washback can be identified. This design can answer the research questions of the researcher; this means that the quantitative longitudinal study can be used to investigate washback.

Similarly, paper 2 chose a two-year longitudinal study, but the researchers used observation which was divided into two parts: baseline observation and observation program. Baseline observation was conducted to observe the learning activities before the new exam. Observation programs were arranged into six rounds to search the effect of the new exam on classroom activities. This observation took place in the same classrooms. (Caruana et al., 2015) said that a longitudinal study using the same sample enables the researcher to gather the changes on individuals over time; therefore, this design allows the researcher to find the effect of washback in classroom activities before and after the new exam. Paper 2 also did an interview to get information concerning teachers and student's responses to the English language test. The researchers can gain information about the existence of washback. They used an observation checklist to check if teachers and students demonstrate differences in their behavior before and after the implementation of the exam. They analyzed the differences quantitatively. They also did the interview with teachers to get complimentary information on why the teachers apply certain strategies in their teaching activities before and after the new exam. This means that researchers of paper 2 can answer the research question about the effect of testing in the teaching and learning process through qualitative and quantitative approaches. Overall, either a qualitative or quantitative longitudinal study is practical and suits to answer the causal research questions; both designs are valid to this inquiry.

The third analysis is based on the research method. A research method is how the researchers do the research based on the research design (Scholtz, de Klerk, and de Beer, 2020). Research method evaluation is divided into three subsections: participants, data collection, and data analysis. Paper 1 surveyed Hong Kong secondary schools with 35 schools from a 323-school population; this sample represented 11%. The number of participants was 844 in 1994 and 443 in 1995. Although the sample of this study decreased, according to Taherdoost (2016), survey research at least has 100 participants. This study has more than 100 indicating that this study covered a large number of participants. Quantitative research emphasizes a high number of participants for the sake of generalizability which leads to the validity of the research (Yilmaz, 2013); the findings of this research might be helpful to understand the behavior of the population of this study (Cohen, Manion and Morrison, 2013), secondary school students of Hong Kong. With this high number of participants, this quantitative study is advantageous as it can ensure the validity of the research.

Meanwhile, study 2 initially involved 49 secondary schools from 7 regions in observation, but since there was a social disturbance, only data of 18 schools were analyzed. This qualitative study also involved many sites, 49 and 18 schools which can contribute to external validity (Hayashi, Abib and Hoppen, 2019) or transferability. Transferability means that the study's findings can be generalized. Besides multi-site research, a thick description of the context of the study also contributes to transferability. A thick description of qualitative research allows the study to be generalized to other typical contexts (Cohen et al., 2013). Study 2 provided a detailed description of the research context: the educational context in Sri Lanka, the O-level Language exam and the textbook, and the process of project O-level exam. This thick description can contribute to the validity of the research. To conclude, each approach: quantitative and qualitative, has its strength to increase the validity of research in terms of sampling. Quantitative study with a high number of participants and qualitative approach with multiple sites and thick description of research context make both methods suitable and useful to investigate the effect of exam in language classroom activities.

In terms of data collection, study 1 collected data using a five-point Likert scale survey of students' perception of classroom activities during the old and new exams. The survey items were derived from multiple sources: reports of the Hong Kong government, past literature, teachers' and students' interviews, and sets of questions related to teacher's perspectives. In this way, the researcher has ensured the instrument to represent the construct of students' perception to washback in language exam, enhancing the content validity of measure; De Vaus (2013) claimed that content validity of measure relates to the comprehensiveness to cover the domain. Furthermore, the researcher developed some indicators for detail on the struct of student's perceptions. In detail, to measure students' attitudes toward teaching, four indicators were used. Meaning, the researcher had maintained the reliability of the survey set. This aligns with De Vaus' (2013) suggestion that a reliable scale consists of a group of questions representing a concept. Moreover, the pilot survey was conducted on two old and new exam cohorts in 1994 and 1995. Pilot testing is an effective way to eliminate ineffective questions in surveys to increase the validity and reliability of the analyzed (Vaus, 2013). Also, the researcher analyzed the data using the SPSS program, which other researchers can replicate. Quantitative study leads the research constructed in a systematic procedure to develop the reliability and validity of research to collect data. In this way, this kin results approach might contribute to objective results representing students' opinion toward the effect of new exams in the classroom; these validity and reliability procedures of data collection contribute to using a quantitative approach in the washback study.

Concerning the ethical issue of data collection through a questionnaire, some points related to protection to participants should be considered: informed consent, participants' rights, issue of beneficence, and confidentiality (Cohen et al., 2013). Paper 1 did not explicitly mention informed consent, but it stated that the number of participants in this study decreased, indicating that participants had the right to withdraw from the study. Some questions were also not answered; meaning, they were not forced to answer all of the questions. The report did not reveal participants' identities; this means that the study kept the confidentiality of participants. Furthermore, all items in the questionnaires in this study did not contain sensitive questions. Also, this research is useful to improve the quality of classroom activities that contribute to participants' learning; this study has fulfilled the ethical issue about the beneficence of research to participants. Meanwhile, Elo et al. (2014) used trustworthiness for a qualitative study to replace the validity and reliability terms. Trustworthiness consisted of credibility, confirmability, transferability, and dependability. In terms of credibility, Lincoln Guba further explained that it can be achieved through triangulation with multiple sources of information, multiple methods, and multiple inquirers.

Concerning study 2, it had some data collections such as observation and interview. The researchers gathered data from various ways which can improve the credibility of finding and this is important to know the phenomenon of washback which is complex. These multiple data collections facilitate researchers to get an in-depth understanding of washback. Furthermore, various methods led to numerous sources of information. The researcher described classroom activities based on observation and why the teachers do specific actions in the classroom through interviews. These multiple sources will increase the credibility of the finding. Also, the researchers of study 2 involved seven observers in which it will be beneficial to confirm conclusions from different perspectives of observers at other points of time and place. Data collection in a qualitative approach through an interview can be useful to search the effect of language exams and to increase the validity or trustworthiness of the research. Overall, both methods have ways to secure the validity and reliability of research findings, albeit in different ways; so, data collection in either quantitative or qualitative approach can be used to gather information of washback.

With regards to ethical issues in observation, there is a dilemma of overt and covert observation. In overt observation, the subjects know that they are observed, so they are aware that they are part of the research. However, this might decrease the study's validity; students know the observer's existence so that they might change their behavior (Cohen et al., 2013). However, this might be a dilemma; participants do not know

that they are observed in the covert observation. For example, students do not realize that they are observed using CCTV. This will increase the validity of research as students will perform their usual behavior; however, it will raise ethical concerns that can violate participants' rights. About study 2, the researchers employed overt observation; students and teachers knew the purpose of observers in the classroom. Although this is overt observation, the observer did 6 rounds of observation that can decrease students' tension as they have been familiar with the observers. So, the students and teachers might perform their actual behavior in the classroom.

The last evaluation is data analysis. It is essential to know how both approaches analyzed the data to investigate washback in language testing better. Study 1 used a quantitative approach with quantitative data, so the finding of this study was inferred from statistic calculation (Muijs, 2013). First, a comparative summary for each classroom activity between the year 1994 (cohort 1) and 1995 (cohort 2) was displayed. Then, the researcher inferred the significance of activity that occurred mainly during the implementation of the new exam. The findings were that the activities such as speaking, carrying out discussion, and playing games were more often in 1995 than in 1994. These activities were in line with the format of the new English exam that was communicative-based; this indicated positive direct washback. The researcher of study 1 concluded that the communicative classroom activities were the effects of the new exam that adopted the communication approach. However, this conclusion can be misleading as there might be another factor influencing teachers in teaching strategy. For example, the textbook that contains more practices relating to communication might lead teachers to teach communicatively; in other words, the teaching strategies might adopt books instead of testing. Therefore, the subsequent investigation is vital to know why the teachers do so, but in a quantitative approach, the interview process asking people's opinions might be regarded as the researcher's intervention to the study's finding. This leads to the drawback of the quantitative approach in washback research.

Meanwhile, paper 2 adopted a mixed-method: a qualitative study using interviews and a quantitative approach using checklist observation. The data then was analyzed using SPSS-X to code the result of checklist observation. Then, the researchers drew a conclusion based on the data. The researchers found that based on the observation, there was no washback in terms of teaching methodology because teachers taught in the same way during baseline observation and observation program. However, after the observation, the researcher did the interview. The data of the interview were coded and classified.

The result was surprising that the teachers had no idea about the exam form and claimed that they never attended the training sessions to explain what kind of skills were assessed in the exam. So, they felt unable to prepare students for the exam and felt willing to prepare students for the exam if they knew what the exam was like. It can be inferred that the washback might occur if the teachers know the format of the exam. Therefore, a clear explanation can be achieved after further investigation.

It can be concluded that the causation investigated with quantitative only answers the hypothesis of cause and effect, but it does not tell in detail why it happened. It is in line with Muijs' (2013) idea that the quantitative approach produces superficial findings. It cannot explore the problem like what the qualitative approach offers with interview and observation in study 2. The researcher can find the answer to how the process occurred and the reason for the phenomenon (Vasquez and Stensland, 2016). For example, in paper 2, the researcher found that the exam itself cannot affect teaching and learning. Still, there might be other factors, such as teacher's training, lack of test material, lack of communication between test designers and teachers, and lack of understanding to test. This detailed information can be accessed only through observation and interview. Cohen et al. (2013) commented that observation could not give a whole picture of what happened in the classroom. Therefore, the observation should be followed up with direct access to teachers and students to get the reasons for classroom activities chosen in the classroom washback study. In conclusion, by using a qualitative approach with multiple methods, the researchers might find what happened and how and why that happened (Vasquez and Stensland, 2016). However, the quantitative approach may excel in the numerous

participants in the study (Prasetyo, Kristin, and Anugraheni, 2019) like what has been displayed by paper 1. This means that the results of washback research might be generalized in another context. Therefore, the writer of this paper recommends that mixed-method research might be better to investigate the complexity of the phenomenon of washback as it employs multi-methods to get access to in-depth understanding through qualitative study and get research findings with numerous data using statistical analysis through quantitative approach so that the study is objective and can be generalized in other similar contexts.

CONCLUSION

This paper analyzed the contribution of quantitative and qualitative approaches for exploring washback in English language testing. This researcher concludes that each approach contributes to washback research. Quantitative may contribute to the objectivity of research for washback and may generalize findings due to the involvement of numerous participants. However, washback researchers may not gain in-depth information on why teachers and students have particular behavior in response to English language testing. Therefore, as each approach has its strength, this researcher recommends that washback researchers use mixed-method studies for investigating washback in English language testing.

REFERENCES

Abu-alhaija, A. S. (2019). From epistemology to Structural Equation Modeling: An essential guide in understanding the principles of research philosophy in Selecting the Appropriate Methodology. *Australian Journal of Basic and Applied Sciences*. <https://doi.org/10.22587/ajbas.2019.13.9.12>

Ali, M. M., & Hamid, M. O. (2020). Teaching English to the test: Why does negative washback exist within secondary education in Bangladesh? *Language Assessment Quarterly*, 17(2), 129–146. <https://doi.org/10.1080/15434303.2020.1717495>

Alla Baksh, M. A. K., Mohd Sallehudin, A. A., & Siti Hamin, S. (2019). Examining the factors mediating the intended washback of the English language school-based assessment: Pre-Service ESL teachers' accounts. *Pertanika Journal of Social Sciences and Humanities*, 27(1), 51–68.

Arthur, J. (2012). *Research methods and methodologies in education*. SAGE Publications Inc.

Asma, A., Sabeen, Q., & Isabel, W. (2014). Investigating the washback effect of the Pakistani intermediate English examination. *International Journal of English and Literature*, 5(7), 149–154. <https://doi.org/10.5897/ijel2013.0521>

Austin, Z., & Sutton, J. (2014). Qualitative research: Getting started. *Canadian Journal of Hospital Pharmacy*, 67(6), 436–440. <https://doi.org/10.4212/cjhp.v67i6.1406>

Beikmahdavi, N. (2016). Washback in language testing: Review of related literature first. *International Journal of Modern Language Teaching and Learning*, 1(4), 130–136. www.ijmltl.com

Caruana, E. J., Roman, M., Hernández-Sánchez, J., & Solli, P. (2015). Longitudinal studies. *Journal of Thoracic Disease*, 7(11), E537–E540. <https://doi.org/10.3978/j.issn.2072-1439.2015.10.63>

Cheng, L & Curtis, A. (2016). Washback or backwash: A review of the impact of testing on teaching and learning. In A. Cheng, L, Watanabe, Y. & Curtis (Eds.), *Washback in language testing: Research contexts and methods*. Lawrence Erlbaum Associates, Inc.

Cheng, L. (1998). Impact of a public English examination change on students' perceptions and attitudes toward their English learning. *Studies in Educational Evaluation*, 24(3), 279–301. [https://doi.org/https://doi.org/10.1016/S0191-491X\(98\)00018-2](https://doi.org/https://doi.org/10.1016/S0191-491X(98)00018-2)

Cohen, L., Manion, L. & Morrison, K. (2013). *Research methods in education*. Routledge.

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE Publications Inc.

2309 *The Advantages and Disadvantages of Quantitative and Qualitative Approach for Investigating Washback in English Language Testing* – Venny Karolina, Muaffi Alif, Sudharni Sudharni
DOI: <https://doi.org/10.31004/edukatif.v3i5.530>

De Vaus, D. (2013). *Surveys in social research*. Routledge.

Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). Qualitative Content Analysis. *SAGE Open*, 4(1), 215824401452263. <https://doi.org/10.1177/215824401452263>

Green, A. (2013). Washback in language assessment. *International Journal of English Studies*, 13(2), 39–51. <https://doi.org/10.6018/ijes.13.2.185891>

Hayashi, P., Abib, G., & Hoppen, N. (2019). Validity in qualitative research: A processual approach. *Qualitative Report*, 24(1), 98–112. <https://doi.org/10.46743/2160-3715/2019.3443>

Hung, S. T. A. (2012). A washback study on e-portfolio assessment in an English as a Foreign Language teacher preparation program. *Computer Assisted Language Learning*, 25(1), 21–36. <https://doi.org/10.1080/09588221.2010.551756>

Hung, S. T. A., & Huang, H. T. D. (2019). Standardized proficiency tests in a campus-wide English curriculum: a washback study. *Language Testing in Asia*, 9(1). <https://doi.org/10.1186/s40468-019-0096-5>

Jung, H. (2019). The evolution of social constructivism in political science: Past to present. *SAGE Open*, 9(1). <https://doi.org/10.1177/2158244019832703>

Kivunja, C., & Kuyini, A. B. (2017). Understanding and applying research paradigms in educational contexts. *International Journal of Higher Education*, 6(5), 26. <https://doi.org/10.5430/ijhe.v6n5p26>

Lauer, P. A. (2013). *An education research primer: How to understand, evaluate and use it*. Wiley.

Lichterman, P., & Reed, I. A. (2015). Theory and contrastive explanation in ethnography. *Sociological Methods and Research*, 44(4), 585–635. <https://doi.org/10.1177/0049124114554458>

Lincoln, Y. S., Lynham, S. A., & Guba, E. G. (2011). Paradigmatic controversies, contradictions, and emerging confluences, revisited. In Y. S. Denzin, N. K., & Lincoln (Ed.), *The Sage handbook of qualitative research* (pp. 97–128). SAGE Publications Inc.

Mertens, D. M. (2014). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed method*. SAGE Publications Inc.

Muijs, D. (2013). *Doing quantitative research in education with SPSS*. SAGE Publications Inc.

Mustika, N., & Lestari, R. (2020). Metode Bermain Peran dalam Meningkatkan Kemampuan Berbicara Bahasa Inggris Mahasiswa STIKes Perintis Padang. *Edukatif: Jurnal Ilmu Pendidikan*, 2(2), 202–209. <https://doi.org/10.31004/edukatif.v2i2.125>

Neuman, W. L. (2011). *Social research methods: Qualitative and quantitative approaches*. Pearson.

Palinkas, L. A. (2014). Causality and causal inference in social work: Quantitative and qualitative perspectives. *Research on Social Work Practice*, 24(5), 540–547. [https://doi.org/https://doi.org/10.1177%2F1049731514536056](https://doi.org/10.1177/216049731514536056)

Phelps, R. P. (2016). Teaching to the test: A very large red herring. *Nonpartisan Education Review*, 12(1), 1–17.

Polesel, J., Dulfer, N., & Turnbull, M. (2012). *The experience of Education: The impacts of high stakes testing on school students and their families*. University of Western Sydney.

Prasetyo, W. E., Kristin, F., & Anugraheni, I. (2019). Penerapan Model Pembelajaran Cooperative Learning Tipe Group Investigation Untuk Meningkatkan Kerjasama Dan Hasil Belajar Mapel Ipa Siswa Kelas 4. *Edukatif: Jurnal Ilmu Pendidikan*, 1(2), 64–71. <https://doi.org/10.31004/edukatif.v1i2.7>

Punch, K.F. & Oancea, A. (2014). *Introduction to research methods in education*. SAGE Publications Inc.

Scholtz, S. E., de Klerk, W., & de Beer, L. T. (2020). The use of research methods in psychological research: A systematised review. *Frontiers in Research Metrics and Analytics*, 5, 1–17. <https://doi.org/10.3389/frma.2020.00001>

2310 *The Advantages and Disadvantages of Quantitative and Qualitative Approach for Investigating Washback in English Language Testing* – Venny Karolina, Muaffi Alif, Sudharni Sudharni
DOI: <https://doi.org/10.31004/edukatif.v3i5.530>

Small, M. L. (2013). Causal thinking and ethnographic research. *American J. of Sociology*, 119, 597–601. <https://www.journals.uchicago.edu/doi/abs/10.1086/675893?journalCode=ajs%0A%0A>

Sultana, N. (2018). Investigating the Relationship between Washback and Curriculum Alignment: A Literature Review. *Canadian Journal for New Scholars in Education*, 9(2), 151–158. <https://cjc-rcc.ucalgary.ca/index.php/cjnse/article/view/53107>

Sykes, B., Plümper, T., Troeger, V., & Neumayer, E. (2019). Case selection and causal inferences in qualitative comparative research. *Plos One*, 14(7), 0219727. <https://doi.org/10.7910/DVN/3H5EDP>

Taherdoost, H. (2016). Sampling methods in research methodology ; How to choose a sampling technique *International Journal of Academic Research in Management (IJARM)*, 5(2), 18–27.

Taqizadeh, M., & Birjandi, P. (2015). What drives high school English teachers to teach the way they do? An Investigation of the washback effect of the university entrance examination in Iran. *Applied Linguistics and Language Research*, 2(2), 8–20. <http://jallr.ir/index.php/JALLR/article/view/63>

Tsagari, D. & Cheng, L. (2017). Washback, impact, and consequences revisited. In S. Shohamy, E., Or, L. & May (Ed.), *Language testing and assessment*. Springer Publishing Company.

Vasquez, M., & Stensland, M. (2016). Adopted children with reactive attachment disorder: A qualitative study on family processes. *Clinical Social Work Journal*, 44(3), 319–332. <https://doi.org/10.1007/s10615-015-0560-3>

Wall, D. & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*, 10(1), 41–69. <https://doi.org/https://doi.org/10.1177%2F026553229301000103>

Yilmaz, K. (2013). Comparison of quantitative and qualitative research traditions : Epistemological , theoretical , and methodological differences. *European Journal of Education*, 48(2), 311–325. <https://about.jstor.org/terms>