

English High-stakes Testing in Indonesia: Friend or Foe?

Maya Puspitasari

Universitas Terbuka

maya_p@ecampus.ut.ac.id

Muhammad Arifin Pelawi

Influx Indonesia

ari.pelawi@influx.com

Corresponding email: maya_p@ecampus.ut.ac.id

Abstract

The Education Ministry has eliminated standardised high-stakes testing called Ujian Nasional (UN) since 2021 throughout schools all around Indonesia. However, it might be practical to look backwards at the UN and its implementation with the impact it gives on the teachers. This article has a concern about the washback consequences of a high-stakes exam on teachers. The effects were investigated within the specific context of the 2016 UN for students in Indonesia at the junior secondary level. A semi-structured interview with eight English teachers was set up to explore the teachers' perceptions towards the UN. Regarding feelings, all participants shared divergent feelings ranging from disinterested to excessive levels of fear or anxiety and mixed feelings. The data also showed mixed opinions, both positive and negative, regarding the examination. The students were thought to be motivated to study because the examination was used for selection purposes. The results did, however, also imply that the students' performance on the national exam was thought to have been assessed unfairly. The instructors demonstrated how the exam affected their teaching roles and methods. Making the national exam a low-stakes test was one of the suggestions made by attendees to enhance it.

Keywords: junior high school; standardised testing; teachers; washback

Introduction,

Indonesia had implemented UNs with different names since the 1950s to measure the student's achievement at school. This standardised testing had been administered nationally for different purposes including summative and achievement purposes. The Indonesian government considered the *Ujian Nasional* (UN) as an efficient tool to evaluate the country's educational system. It is in line with Au's (2021) proposal. He argued that standardised testing is "the perfect technology for the efficient categorizing, sorting, and ranking of human populations" (p. 64). Since Indonesia has huge numbers of the population, standardised testing could play a significant role and be used as part of decision-making to determine which

education policy should be implemented or amended. It was likely the same motive because the Indonesian government chose the UN as Indonesia has huge numbers of population.

Since the outcomes of the national exam are used to evaluate the performance of the students, it might lead teachers to be under the pressure of accountability (Hofflinger & von Hippel, 2018; Holloway and Brass, 2018; Vasquez Heilig, Brewer and Ojeda, 2018; Adler-Greene, 2019; Perryman and Calvert, 2020; Asian Development Bank (ADB) and Organization for Economic Cooperation and Development (OECD) (2015: 314) noted that one of the main causes of systemic cheating and bad administration, as evidenced by the inflated mean score of some cities, councils and provinces, is the concern that results from UN will be used to evaluate institutions of learning, cities, councils or provinces. It then possibly puts the teachers to concentrate on the subjects that would be tested in the UN and this is how the washback effect might happen.

There are numerous definitions for the word "washback" (Alderson, 1984; Hamp-Lyons, 1998; Spratt, 2005; Tsagari, 2011; Wall, 2013; Finardi and Archanjo, 2018; Dong, 2020; Liu and Yu, 2021). The term itself has been initially introduced about a few decades ago and has been developed continuously. Alderson (1984) is among those first who attempted to define the term by describing it as an impact of a high-stakes test on syllabus and teaching. Shohamy (1998) a decade later developed the idea of washback as a major impact of a language test on a teacher's instruction. Classroom practices before students sit for a high-stakes test might also be dominated by the exam preparation cycle. Such preparation is including "a teacher tailors classroom practices to meet the test requirements and improve students' scores by providing techniques to eliminate the distractors—wrong answers in the multiple-choice questions—and to do well on the test" (Imsa-ard, 2021: 61).

The UN took in the final year of schooling (sixth grade for primary school students and third grade for junior and senior secondary students) was used to measure the students' attainment in every school and a pass entrance for the next schooling level. Kirkland (1971: 307) notes:

"Test effects are understandably most pronounced and most potent on the individual being tested. These effects are complicated and interwoven with effects on parents, teachers, schools, and society, and it is difficult to separately consider any one of these sectors."

In the literature from all over the globe, there has been evidence of a significant connection between high-stakes exams and classroom instruction. (Counsell and Wright, 2018; Veselak, 2018; Dong, 2020, Tan, 2020). Given that teachers seem to be important facilitators in pupils' classroom learning experiences (Kohler, 2015; Lai, et.al., 2016; Keiler, 2018), there is a tendency the teachers may also help their pupils to beat the test. Au (2023: 72) argued that when:

"a commitment to the assumptions of a standardized test as objective is made, students are viewed as commodities. This is evident in the reduction of students (or teachers or schools) to mere test scores, where their value in the educational marketplace is measured by their testing achievements."

From different contexts and backgrounds, some evidence has suggested that high-stake testing has washback effects (negatively or positively) on teachers and teaching practices. A clear example of possible washback to the students is found in Zhan and Wan's (2016) study. In Guangdong Province, mainland China, they interviewed 20 students at a high school in Guangzhou. In preparing for the Computer-based English Listening and Speaking Test (CELST), the students focused on gaining high scores on the test by using past and mock test papers as the preparation materials. Another study conducted by Ali and Hamid (2020) with twenty English teachers in Chittagong, Bangladesh shows that their participants perceived public examinations had a negative effect on pedagogy. The participating teachers held the view that pupil performance on the public examination determined whether educational practises were successful or unsuccessful.

The previous studies have highlighted the existence of washback of high stakes testing from various schooling level and countries. A little is known about research studies embedding feelings and teaching focus of teachers regarding the national examination in Indonesian context. The present study tries to meet the gap by investigating the washback effect of the national examination on teachers. A significant addition to the body of knowledge on the high-stakes test on teachers and teaching practises could then be made by researching how teachers in Indonesia view the national test and its washback effect. The purpose of this study is to respond to the research queries: 1) How do Indonesian teachers feel about the UN's effect? 2) How does the UN have an impact on the classroom teaching methods used in final-year classes that follow the UN?

Research Methodology,**Context of the Study**

Standardised testing had been implemented nationally into the Indonesian educational system since the 1950s. Since 2005 the Indonesian Ministry of Education and Culture had administered standardised testing called *Ujian Nasional* (UN) to measure the excellence of education within the nation. This high-stakes test had achievement purposes for students at grade primary school (grade 6), an intermediate school (grade 9) and a high school (grade 12). For ninth-grade students, there were four subjects tested: science, mathematics, Indonesian and English. This study will discuss the English UN and exclude the other subjects tested. Fifty multiple-choice questions in English should be answered by the students in 120 minutes. The questions covered reading, listening and grammar.

Research Participants

In qualitative research, the goal is to create a thorough investigation of a subject matter rather than to generalise to a population (Creswell, 2012: 206). Understanding how teachers interact with the national exam's authority and its impact on their practises and perceptions was crucial. The experiences that teachers have had in their lives may have an impact on how they perceive the washback effect of the standardized test. Due to the limitation of time, funds and access, a smaller group; Kabupaten Bandung, a city in West Java was chosen, to gain information from the whole population. There were 227 private junior secondary schools and 75 public junior secondary schools in Bandung during the period of data collection (February to May 2016), which offered young students who had completed primary school a continuation of their education.

The participants were purposefully chosen, ensuring that the data represented people with relevant expertise. Participants were chosen based on their willingness to engage, a mix of gender, the kinds of schools they attended, and their range of teaching experiences. The research involved eight teachers of English subjects from eight intermediate schools. The study's goal was to learn more about how they perceived and dealt with teaching pupils who would take the national test. The participants (six women and two men), who volunteered in semi-structured interviews, shared their perspectives on how the UN affects their classroom management.

Data gathering tool and data analysis

The interviews were mainly guided by the research questions and held in Indonesian for the participants to feel more comfortable expressing themselves. It took up to 60 minutes for each participating teacher and it was video recorded to capture the teachers' facial expressions. An information sheet and a consent letter were given to each participant.

Researchers may be like explorers of new terrain lacking an easy-to-understand chart when analysing qualitative data. (Silverman, 2015: 208). The interview data were then transcribed into Indonesian and translated into English. Pseudonyms were used throughout the transcripts to ensure participants' anonymity. Then, the translated transcriptions were coded by utilizing phrases related to the research questions.

Findings and Discussion**Findings**

Two themes were taken from the analysis of the interview data: teachers' feelings and teaching focus based on the two research questions of this study: 1) How do Indonesian teachers feel about the UN's effect? 2) How does the UN have an impact on the classroom teaching methods used in final-year classes that follow the UN?

Teachers' Feelings

Seven of the participants agreed that the third or final year was the toughest time for the students. Indri, for example, said: "The ninth grade. Since the ninth grade will have the graduation. Therefore, they need to be prepared from the beginning; their graduation, the strategy to give them marks." Reasons for the view that these students were the trickiest year group to teach were likely based on three beliefs. First, Fera, the teacher in an elite private school, believed that the students she taught were very demanding. She said that the students "keep complaining if there is something that doesn't match with what I've explained before. Then I have to explain it one more time." On the other hand, Husna and Mahmud provided a different answer, proposing that classroom supervision was difficult in the final grade because the students were more prone to misbehave. Husna believed that "the ninth grade is the moment when children transform into teenagers," and that as a

result of this change, students now acquire "juvenile delinquency; they start to misbehave" (Mahmud).

The third reason that three respondents agreed to is the UN. Although the UN is not the sole determinant of students' graduation, the teachers made specific it is still affecting the need of students and is affecting the demand for their teaching. The challenge that the students had, said Putri, "they should be well prepared mentally and academically to sit in the UN." Thus, they argued, the situation made additional demands on the teachers to provide students with 'preparation' and 'strategy' to gain good scores (Indri). They believed that the burden was heavier as they are 'the teachers of whose subjects are tested in the UN (are) obliged to graduate their students a hundred per cent no matter what, that is the burden as it is'. From teachers' perspectives, the biggest burden of their teaching was therefore not to educate the students but to graduate and gain good marks. The burden teachers in the study expressed negativity and worry regarding the UN (Fera, Mahmud and Putri), not being transparent on grading (Endang), psychological (Susan), fear (Indri and Husna), and a strong sense of responsibility (Nopi).

Meanwhile, one of the outlined purposes of the UN is to evaluate teaching practices. Six teachers shared similar opinions regarding this issue. Susan, for example, said:

"When I looked at the result of last year's UN, I could recall the memory. I think the way I teach must always be reformed. For example, the method of studying may be altered".

As the teachers evaluate the result with what they need to do in the future, it would mean the focus on what they teach for future students is also affected. The choice of the vocation of the teacher is related to one's family background; six out of eight teachers in this study stated that someone in their family, who worked as a teacher, motivated their career interest in teaching. Furthermore, half of the teachers interviewed worked alongside family colleagues at a school. Husna and Endang confirmed that they were recruited to teach by a family member, enabling them to access the vocation more easily:

"When I was in the third semester at university, my brother became the school principal at this school; I was often offered to teach there, but I always refused. But then my brother asked his wife to convince me... I couldn't refuse" (Husna).

"a relative asked me to become part of the administration staff at this school. I first studied at the university, and then I became a teacher" (Endang).

High expectations in teachers' lives are likely related to the quality of relations between them as individuals and their families as such relationships are central to life in Indonesia. Having family who pursues a similar profession, and/or working with family members at the same school indicates that the level of anxiety of teachers could be higher than those whose family had no teaching background nor worked at the same institution. This was suggested by Nopi, who stated, "I would feel ashamed if later there's even one student that doesn't graduate; I would be ashamed, as I would have failed as a teacher... I'm afraid that the student's score will be low." Working at a school where most of her colleagues were members of her own family increased Nopi's sense of anxiety. Similarly, to Nopi, Putri possessed many family members who worked as teachers, though she did not work alongside any family members. At the start of the interview, she explained how her choice to become an educator had been influenced by her family's professional background:

"My great-grandfather was a school principal, and a teacher before that; my grandfather was also a school principal and a teacher. My uncle was a teacher too. They are from my father's side of the family. On my mother's side, some became teachers and school principals as well... So, the blood to become a teacher had been long rooted".

The perceptions of teachers towards the UN were also found to be inextricably intertwined with their families. Having a family who also worked in the educational field tended for the teachers to perceive a good grade in the UN as an important objective. The negative feelings and attitudes that some participants showed towards the UN appeared to come from the low self-esteem that possibly occurred from a bad result of their students connected to having family members working at the same school.

Achievements achieved by students in the UN could be one of the most important indicators of teaching English to final-year students. Along with expressing her concern for the upcoming national exam, Putri subsequently made an intriguing point regarding her choice:

"I told them if they get [a score of] 10, I will reward them, similarly, to last year when there were two students who got a perfect score in the English test... I told them, 'if there are 10 students who achieve a perfect score, I will award those 10 people as promised'".

The status of the teacher within the wider community was also found to have a considerable influence on motivations for becoming a teacher. Mahmud provided his reasons for becoming a teacher:

"...though his or her wage is low, society will treat him or her more respectfully as a teacher. Their salary is very low compared to those who work in a factory, but people still give more respect to the teachers".

The answers of Nopi, Putri, and Mahmud might also point to a connection between family members' occupations and backgrounds, and social standing, as well as why teachers viewed the UN as a demanding test. It was thought that having students who did well on the UN reflected effective teaching and/or schooling. Thus, the UN's outcome evolved into the primary objective and source of inspiration for instructors of final-year students.

Teaching Focus

Teaching focus includes the changes made within the classroom as a direct result of the UNs. More than half of the teachers shared their perspectives on how the UN-led them to change classroom practices. The teachers were expected to adhere to the lesson plans created for the 2013 curriculum. However, because the textbook's description of the syllabus included material that was not assessed by the UN, they often chose to neglect these parts of the syllabus, instead referring teaching content to the graduate's competency standard (SKL). Putri explained:

"About preparation for the UN, the first thing we look at is the guidance; which material will appear in the UN; which SKL will need to be delivered".

Nopi echoed Putri's statement, noting that she also referred to the SKL as her direction in educating students for the last semester. Nopi disregarded the syllabus's instructions for educating her pupils in the second term because the syllabus's instructions "had been given in the eighth grade." Nopi liked using SKL because the syllabus expands the areas covered by the English instruction, while SKL was more practical and make "the children can get used to" the content that would be tested in the UN. Additionally, the institution where Fera taught based its primary objective of educating ninth-year students on its motto, "succeeding in the UN". Thus, in the last term, Fera only focused on supervising students in preparation for the UN. She also "skip the speaking skill," because it would not be assessed in the English UN. Susan also undertook similar activities in the second term, when the UN became the main focus in teaching final-year students. Therefore, most of the teachers in this research

believed that the UN exerted a significant influence on the practice of teaching and the student's learning.

Most of the teachers held an extra class out-of-school formal hours focusing on exam practise exercises, as exemplified by Nopi, who said, "the school's hours have been extended; there are enhancement hours." Most teachers said that the enhancement classes needed to intensify exam exercises in the second term. It was started in January until the time of UN would be administered in May. This intensive test-practice exercise was undertaken "before the school bell rang...from 6 to 7 a.m." as Husna noted, or after formal school time was ended as Mahmud stated. Meanwhile, Susan and Endang informed us that enhancement classes were practised in their classroom during formal school time. Susan stated that she only delivered the syllabus in the second term within two months while after that they undertook exam practise exercises.

Teachers' devotion to more exam preparation time likely increased as the date for UN administration approached. As the UN was held after two weeks of the school-based test, Endang, for example, used the time slot for the enhancement classes. He stated:

"We think that if time is wasted, it's such a shame. The school principal planned to use it [the two-week slot] for the ninth graders to learn the four subjects. So, within two weeks we can discuss the questions of the past UN".

After the school examination, Endang believed that students only needed to learn the four subjects assessed in the UN and nothing else. To effectively prepare students for the UN, he and other teachers whose courses were also being tested, used all of the class time. Additionally, Endang mentioned that pupils had to take mock exams prior to the UN. The students must do three mock exams organized by "the school, local government and provincial government."

However, a teacher did not agree with enhancement classes outside formal hours. She shared her concerns:

"It's not because I don't need it (the enhancement), but we don't have the time. It's different with the other schools whose students go home at 1 or 2 p.m. Here, the students go home at 3 p.m. So, if we hold enhancement classes, they will get home too late".

Similarly, Fera's schooling lasted for eight hours too. She also did not state the possibility of putting in additional time outside of her teaching responsibilities. Her

teaching focus in the classroom formal time that intensified on coaching for how to beat the UN could be why she had no such additional time for enhancement class.

All teachers in this study also reported that the UN had an impact on the course materials for the second semester during class time and/or in exam practise class. Exam practise questions and previous test papers were provided as materials. Indri's declaration that the UN did not affect her teaching content contradicted the noted materials she sometimes used in the classroom, as she said:

"I take examples from previous questions of the UN...I explain it based on the syllabus; in the task part, I took it from the questions of the previous UN".

Teachers later reported using specific methods where they focused their teaching practices mainly on exam practice exercises during class time. Regarding this, Fera described her strategy that focused on coaching the students, as such materials were believed as essential for her student's success in the UN:

"I only teach them how to do the UN. Then we evaluate who still doesn't understand the questions. I spend most of my teaching time in discussions with the students; to discuss what kinds of questions [covered in the UN] they think are still difficult for them to understand".

One noteworthy situation was found in the statements received from Susan and Putri. They gave some essential steps on how to treat the answer paper. Putri told the students, "Not to blacken the answer sheets first... To mark the multiple choices on the questions," while Susan informed the students, "not to read the whole text as the time will run out [before they finish]." Putri reminded her students that the answer paper could not be folded, torn, dirty or perforated because it would affect negatively the grading as the computer system would not be able to do it correctly. Meanwhile, Susan explained the motivation behind her decision to implement strategic instruction in her classroom, as, "the average score of our students couldn't reach the highest score... I just want to make them get used to the questions of the UN".

Discussion

In straightforward terms, when schools are made accountable for the outcomes of their students, it is asserted, would in turn be transferred to teachers and principals to improve them. It was clear from all the teachers' reports that pressure from the need of school is also involved in influencing teachers' decisions and therefore has a significant effect on the teachers' feelings and what takes place in the classroom. All the teachers' responses reflected that the results of the UN

became the cornerstone in shaping the quality of the school in the community. Their views indicate that there is a belief among stakeholders in a robust connection between the UN and the quality of schools and the effectiveness of teachers. Falabella (2014) described it clearly that testing has shifted from a means to collect information from the students to measure teachers' and schools' accountability.

High-stakes testing as the English high-stakes test in Indonesia could influence teachers and students (test-takers) and teachers (those who have an important role in preparing students for the test). However, the teachers are not the only stakeholders who are affected, as they also interact with school principals and parents within both a school and home context. First, teachers and school principals tended to experience greater pressure and anxiety given that their status can be established or broken depending on the outcomes of the UN. Therefore, teachers and principals also have a responsibility to boost the scores of their students in the UN. Second, students may be anxious, as the grades of the UN substantially impact their opportunities in the future – particularly for the next level of schooling admission. Third, parents' concerns regarding the results of the UN are likely to affect their involvement with their children's learning.

Indeed, having an examination preparation cycle in the classroom before the exam begins would be advantageous for a test taker to get a good result. But a question should be asked further. The grades achieved by the students in the UN might be boosted, but does it mean the learning quality of the students is as high as the scores tell? The assessment system is supposed to help to learn in the classroom. Indonesia has always said in the formal text that assessment is a small part of learning but contrary to the 'concept', the small "reality" shown in this study shows that it has become "bigger" than the learning itself, as learning only acted as an action to pass the test. These movements have neglected that teaching, in the deepest and most basic sense of the meaning, is a commitment to help students make learning humanise the human being. Holloway and Brass (2017: 380) concluded that might be:

"a shift in governmentality where objectification, quantification, and measurement are no longer treated as antithetical to teacher professionalism, but as precisely what teachers need to know and monitor themselves, improve themselves, and fashion themselves as professionals. The distance between the teachers and the accountability apparatus has all but collapsed, fabricating a new kind of teacher,

whose value is oriented to markets, management, and numerical performance indicators".

In this article, we want to bring an exploratory study to urge for making assessment conceptions more compatible with commitments of teaching. We do not look to provide an "irrefutable fact" that high-stakes tests are really "bad" for teaching practice, however, we hope that more research can be provided about how "a success story" of school in assessment records is achieved. This is on the one hand to understand the struggle over assessments' meaning and concepts to teachers. On the other hand, we need to look for a way to make quality mechanisms assessment re-signified, re-imagined, and transformed to suit the need of teaching and learning processes, not vice versa.

Conclusion

To understand the teachers' perspectives on high-stakes UNs for intermediate schools in Indonesia, an interpretive theoretical perspective was applied. What became transparent from interviews with the teachers was that the grades of the UN were perceived by the teachers and other stakeholders as a tool to evaluate the process of learning and teaching. The teachers gave firm views of what they and the school principals considered to be important goals for final-year students. It then influenced the teaching focus in the classroom. They likely preferred to teach to the test rather than teach to gain knowledge or implement deep learning. Conducting qualitative approach has a potential bias as the results could lead to be subjective. Further research inquiries are recommended to reduce bias by studying larger group of participants and different settings. The national examination has been abolished by the Indonesian Ministry of Education and Culture since a few years ago. It might also be interesting to know if the teachers' anxiety has been shifted from high to low regarding to the policy.

Acknowledgement

This article is part of the first author's doctoral dissertation supervised by Professor E. Louise Hayward and Professor Oscar Odena from the School of Education, University of Glasgow. The research project was fully funded by the Indonesian Ministry of Education and Culture.

References

- Brown, J.D. (1995). *The elements of language curriculum: A systematic approach to program development*. Boston: Heinle & Heinle Publisher
- Adler-Greene, L. (2019). Every Student Succeeds Act: Are schools making sure every student succeeds. *Touro L. Rev.*, 35, 11.
- Alderson, J. C. (1984). Testing, the teacher and the student. *Cahiers de l'APLIUT*, 3(3), 7-21. <https://doi.org/10.3406/apliu.1984.1788>
- Alderson, J. C., & Wall, D. (1993). Does washback exist?. *Applied linguistics*, 14(2), 115-129. <https://doi.org/10.1093/applin/14.2.115>
- 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>
- Andrews, S. (2004). Washback and curriculum innovation. In *Washback in language testing* (pp. 59-72). Routledge.
- Asian Development Bank (ADB) & Organisation for Economic Co-operation and Development (OECD). (2015). *Education in Indonesia: Rising to the challenge*. Paris: OECD Publishing.
- Au, W. (2023). *Unequal by design: High-stakes testing and the standardization of Inequality* (2nd ed.). New York: Routledge.
- Counsell, S. L., & Wright, B. L. (2018). High-stakes accountability systems: Creating cultures of fear. *Global Education Review*, 5(2), 189-202.
- Dong, M. (2020). Structural relationship between learners' perceptions of a test, learning practices, and learning outcomes: A study on the washback mechanism of a high-stakes test. *Studies in Educational Evaluation*, 64, 100824. <https://doi.org/10.1016/j.stueduc.2019.100824>
- Falabella, A. (2014). The Performing School: The Effects of Market & Accountability policies. *Education policy analysis archives*, 22(70). <https://doi.org/10.14507/epaa.v22n70.2014>
- Finardi, K., & Archanjo, R. (2018). Washback effects of the Science without Borders, English without Borders and Language without Borders programs in Brazilian language policies and rights. In *Language policy and language acquisition planning* (pp. 173-185). Springer, Cham. https://doi.org/10.1007/978-3-319-75963-0_10
- Hofflinger, A., & von Hippel, P. T. (2020). Missing children: how Chilean schools evaded accountability by having low-performing students miss high-stakes tests. *Educational Assessment, Evaluation and Accountability*, 32(2), 127-152. <https://doi.org/10.1007/s11092-020-09318-8>
- Holloway, J., & Brass, J. (2018). Making accountable teachers: The terrors and pleasures of performativity. *Journal of education policy*, 33(3), 361-382. <https://doi.org/10.1080/02680939.2017.1372636>
- Imsa-ard, P. (2021). Choices of teaching practices: Does the English UN cause Thai EFL teachers to teach to the test. *TESOL International Journal*, 16(5), 56-86.
- Keiler, L. S. (2018). Teachers' roles and identities in student-centered classrooms. *International journal of STEM education*, 5(1), 1-20. <https://doi.org/10.1186/s40594-018-0131-6>

- Kirkland, M. C. (1971). The effects of tests on students and schools. *Review of Educational Research*, 41(4), 303-350. <https://doi.org/10.3102/00346543041004303>
- Lai, C., Yeung, Y., & Hu, J. (2016). University student and teacher perceptions of teacher roles in promoting autonomous language learning with technology outside the classroom. *Computer Assisted Language Learning*, 29(4), 703-723. <https://doi.org/10.1080/09588221.2015.1016441>
- Liu, X., & Yu, J. (2021). Relationships between learning motivations and practices as influenced by a high-stakes language test: The mechanism of washback on learning. *Studies in Educational Evaluation*, 68, 100967. <https://doi.org/10.1016/j.stueduc.2020.100967>
- Perryman, J., & Calvert, G. (2020). What motivates people to teach, and why do they leave? Accountability, performativity and teacher retention. *British Journal of Educational Studies*, 68(1), 3-23. <https://doi.org/10.1080/00071005.2019.1589417>
- Hamp-Lyons, L. (1998). Ethical test preparation practice: The case of the TOEFL. *Tesol Quarterly*, 32(2), 329-337. <https://doi.org/10.2307/3587587>
- Kohler, M. (2015). Teachers as mediators in the foreign language classroom. In *Teachers as Mediators in the Foreign Language Classroom. Multilingual Matters*. <https://doi.org/10.21832/9781783093076>
- Silverman, D. (2015). *Interpreting qualitative data* (5th ed.). London: SAGE Publications Ltd.
- 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. <https://doi.org/10.1191/1362168805lr152oa>
- Tan, C. (2020). Beyond high-stakes exam: A neo-Confucian educational programme and its contemporary implications. *Educational Philosophy and Theory*, 52(2), 137-148. <https://doi.org/10.1080/00131857.2019.1605901>
- Tsagari, D. (2011). Washback of a high-stakes English exam on teachers' perceptions and practices. *Selected papers on theoretical and applied linguistics*, 19, 431-445. <https://doi.org/10.26262/istal.v19i0.5521>
- Vasquez Heilig, J., Brewer, T. J., & Ojeda Pedraza, J. (2018). Examining the myth of accountability, high-stakes testing, and the achievement gap. *Journal of Family Strengths*, 18(1), 9.
- Veselak, K. M. (2018). Teachers' Beliefs About the Purpose of Education in a High-Stakes Testing Environment. *National Teacher Education Journal*, 11(3).
- Wall, D. (2013). Washback. In *The Routledge handbook of language testing* (pp. 93-106). Routledge.
- Zhan, Y., & Wan, Z. H. (2016). Test takers' beliefs and experiences of a high-stakes computer-based English listening and speaking test. *RELC Journal*, 47(3), 363-376. <https://doi.org/10.1177/0033688216631174>