Exploring Saudi Students' Perceptions of National Exams: A Washback Study

Ahmad ALSAGOAFI¹

¹King Faisal University, Saudi Arabia,
aalsagoafi@kfu.edu.sa

Abstract: Empirical research on washback and impact of tests (either positive or negative) on teaching and learning in the language assessment field has proliferated in approximately the last three decades since the publication of Alderson and Wall's (1993) seminal paper. The objective of this study is to explore the positive washback of the Scholastic Achievement Admission Test (SAAT) and the General Aptitude Test (GAT) from the perceptions of secondary school students in Saudi Arabia since this topic is underresearched in the literature. While the existing research has provided insight into the validity and reliability of these tests, we have a limited understanding of how these tests can positively impact the performance of students.

This study employed a qualitative methodology using case study as a research design. Data was collected from 103 participants through the means of semi-structured questionnaires and semi-structured interviews. To interpret the findings, a content analysis was carried out through NVivo software. The analysis revealed that the participants were positively impacted by the test. This influence was manifested in students' low pressure test-taking experiences, their successful honing of soft skills needed for future university study instead of reliance merely on indoctrination or the memorization of information and the feasibility of enrollment in universities because the Secondary School Certificate (SSC) is no longer the sole indicator of academic performance. This study concluded that students' scores increased due to the method of testing, test preparation and test retaking, which played key roles in promoting score gains and thus accentuating a positive washback.

Keywords: *GAT & SAAT, Positive washback, Soft skills, Case study, High-stakes tests.*

How to cite: Alsagoafi, A. (2021). Exploring Saudi Students' Perceptions of National Exams: A Washback Study. *Revista Romaneasca pentru Educatie Multidimensionala*, 13(1Sup1), 213-234. <https://doi.org/10.18662/rrem/13.1Sup1/393>

Introduction

The term “washback” is deeply rooted in applied linguistics. Washback commonly used in the language assessment literature (see, e.g., Bachman, 1990; Alderson & Wall 1993; Davies et al., 1999; Cheng, Watanabe, & Curtis, 2004; Green, 2007). However, in the field of education, the term “backwash” is used with the same connotations (see, e.g., Gipps, 1994; Broadfoot, 1996; Hughes, 2003). Thus, the two terms are used interchangeably to refer to one concept, namely, the impact of tests on instruction as Davies et al., (1999) refer to it, on teaching as Alderson & Wall (1993) described it or on learning as Hughes (2003) suggested.

There is a further distinction in the field of language assessment between impact and washback. A test impact according to Wall (1997, p. 291) is “any of the effects that a test may have on individuals, policies or practices, within the classroom, the school, the educational system or society as a whole.”. On the other hand, she defined washback as “the effects of tests on teaching and learning”(Wall, 1997, p.291). While these terms may be seen as referring to separate concepts, some language testers still consider washback to be part of impact (see, e.g., Hamp-Lyons, 1997). Additionally, the traditional view of testing has tended to focus on the quality of tests, i.e., their psychometric properties, which can perhaps be controlled during the test development process. However, the testing experience of the candidates has been overlooked, resulting in washback being considered an exercise of power over test takers (Shohamy, 2001). She called for the examination of how tests are used, their importance in the lives of the candidates and their place in society (Shohamy, 2001). Similarly, Kunnan (2004, p. 39) presented a fairness framework that encompassed the entire “system of a testing practice not just the test itself” where he included five test fairness qualities. One of the most important of these qualities was the social consequences in terms of *washback* and *remedies*, for which the concern is about the testing practices in relation to whether they contribute to “societal equity.”

As such, there has been growing interest in conducting empirical washback studies since the publication of Alderson and Wall’s (1993) seminal paper (see, e.g., Cheng, Sun, and Ma (2015) for a review of the washback research literature from 1993 to 2013). Likewise, the special issue on washback that was published by *Language Testing* in 1996 sparked interest in the topic, with a number of empirical studies being published since. Furthermore, Shih (2010) chronologically tracked the growing number of

studies that have shed light on the phenomenon of washback from as early as the late 1960s until after the new millennium.

Due to the nature of this study, the term “washback” will be used. We can have positive washback (see, e.g., Messick, 1996; Davies et al., 1999; Taylor, 2005; Saif, 2006), neutral (see, e.g., Wilkinson, 1968), or negative (see, e.g., Wiseman, 1961; Sinclair, Fox, Moon, & Stock, 1987; Khaniya, 1990; Messick, 1996; Davies et al., 1999). This paper focuses on the positive washback of testing in Saudi Arabia with special reference to the Scholastic Achievement Admission Test (SAAT), General Aptitude Test (GAT), and Secondary School Certificate (SSC). These tests are considered high-stakes tests. They have important consequences in the lives of the candidates, and as such, they have washback (Alderson & Wall, 1993). Deygers, Van den Branden, and Van Gorp (2018, p. 451) asserted that “a score only becomes *real* when it has real-life consequences.”

Given the concerns about test consequences, it seems that positive washback is achieved through the appropriate use and accurate interpretation of test scores. Messick (1996, p. 245) noted that “in language testing..., what matters are not the processes operative in task performance..., but the processes captured in test scoring and interpretation.”

Previous studies commissioned by the National Center for Assessment (NCA) on the GAT (see e.g., Al-Qataee, 2012a March 21, 2012b August 3; Dimitrov, 2013a January 22, 2013b March 12, 2013c April 29, 2013d May 16, 2013e June 30, 2013f August 19, 2013g September 25, 2013h December 13, 2015; Dimitrov & Shamrani, 2015; Dimitrov, Raykov, & Al-Qataee, 2015; Jehangir, 2015, 2016a April 3, 2016b August 14; Lamprianou, 2013a October 11, 2013b December 1, 2014a September 5, 2014b October 4, 2014c November 9, 2014d December 21, 2015a November 21, 2015b December 17; Sideridis, 2014, 2016; Sideridis & Tsaousis, 2013a December 15, 2013b December 30; Sideridis, Tsaousis, & Al-harbi, 2015) and on the SAAT (see e.g., Al-Owidha 2013; Lee & Aluhiedan, 2013; Dimitrov et al., 2015; Lamprianou, 2015a November 21, 2015b December 17) have examined only the psychometric properties of the tests, focusing on the validity or reliability of the tests. However, this focus does not mean that the accuracy of tests is not a significant factor in the quality of tests. An evaluation of a single aspect of the construct validity of a test through washback is not sufficient if we are interested in test use, which should be part of the broader definition of validity (Messick, 1996). Shohamy (2001, p. 5) maintained that the language testing field has recently considered “issues

about the use of tests – i.e., intentions, effects and consequences” to be part and parcel of testing practices.

Similarly, there have been other studies on the GAT and the SAAT that were not commissioned by the NCA. These studies have focused either on the validity or the reliability of the tests alone or in relation to students’ high school or university GPAs (see, e.g., Alshehri, 2007; Alshumrani, 2007; Balkyoor, 2008; Al-Alwan, 2009; Almehrej, 2009; Alshehri, 2010; Siddiek, 2011; Alsehaly, 2012; Alanazi, 2014; Aldahesh, 2015; Alnahdi, 2015; Alnowibet & Ahmad, 2015; Alshemeri, 2015; Hamdan & Al-Hattami, 2016; Sulphrey, Al-Kahtani, & Syed, 2018).

All of the aforementioned studies have examined only the characteristics of the tests or other relevant variables while ignoring the perspectives of stakeholders. Fortunately, some studies on the GAT and SAAT in the literature have examined test use and, therefore, the consequence of the test for test takers (see, e.g., Alshammmary, 2008; Alfraidi, 2012; Alghamdi, 2012; Alghamdi 2017; Al Ruwaili, 2013; Bobsheet & Al-Ghamdi, 2015; Alazizi, 2016). However, these studies did not investigate in-depth the washback of the tests or their positive effects. Additionally, almost all of them were quantitative except for Al Ruwaili’s (2013) study, which involved the use of interviews with a mixed methods approach. Nonetheless, the students’ voices in Al Ruwaili’s study were overlooked, as the interviews were conducted only with teachers, educational leaders, and university lecturers. The present study examines the positive washback of the GAT and SAAT from the perspectives of Saudi students using a qualitative approach, where it fills this gap in the literature. This paper, based on the above review of literature, has addressed the following main research question and two sub-questions:

The main question was as follows:

What are Saudi students’ perceptions of the GAT and SAAT?

The sub-questions were as follows:

- (1) What roles do the GAT and SAAT play, from the perspectives of the students, in generating a positive washback of testing in Saudi Arabia? Why and in what way?
- (2) How are the results of the GAT and SAAT currently interpreted? How do they affect students’ futures?

Background of the study

Standardized Tests in Saudi Arabia

The NCA developed the GAT in 2002 and the SAAT in 2004. The aim was to unify admission criteria across Saudi universities by providing reliable tests that could predict academic success in the target institutions. In addition, the poor quality of teacher-made achievement tests as well as the overreliance on indoctrination and memorization may have led the NCA to reform education through such tests. Critical thinking, problem-solving, logical relations and the application of knowledge are emphasized more than memorization on the GAT and SAAT tests, which has had an impact on the content of the school curricula. Furthermore, the inflation of high school GPAs, which could sometimes be attributed to nepotism or family connections, no longer occurs, as high school graduates who would like to enroll at universities or colleges need to bring evidence of their achievement on the SAAT and GAT.

Candidates who apply to colleges for majors in art need to take only the GAT, whereas candidates who apply to colleges for majors in science need to take both the GAT and the SAAT. The scores from these two tests are added to the SSC grade to produce a composite score. This composite score is the basis for acceptance into universities or colleges.

General Aptitude Test (GAT)

The GAT is a test that measures language and mathematics through the following abilities: problem-solving behavior, reading comprehension, inferential abilities, inductive abilities and logical relations. The test consists of six sections with a total of 120 multiple choice questions (68 verbal and 52 quantitative). The duration of the test is 3 hours.

Scholastic Achievement Admission Test (SAAT)

The SAAT measures students' achievement in their secondary school science subjects, namely, chemistry, physics, biology, and mathematics. The test consists of 120 multiple choice questions distributed as follows: Year (20%), Year (30%) and Year (50%). The duration of the test is 125 minutes.

Secondary School Certificate (SSC)

Prior to the introduction of the abovementioned tests, there was a public examination administered in the last year of secondary school. Admission into universities was granted to those students who achieved the

passing grade for each university. Criteria for admission varied across colleges and universities. Students were under great pressure as a result of these public high stakes examinations. Parents were worried that their children would not be able to enroll at university. Teachers were eager to determine the best practices to prepare their students. This form of public examination may have caused negative washback. At the time, the Ministry of Education needed a standardized test similar to the Scholastic Achievement Test (SAT) in the United States. The intention may have been to lessen the negative washback of public examinations, such as that evidenced in the poem above. The introduction of the GAT and SAAT was thus aimed to encourage positive washback. This study has set out to further explore the positive washback of these tests based on students' perceptions.

Methods

In this study, a case study approach was used to explore the positive washback of the GAT and SAAT. The investigation of factors is embedded in real-life situations, and phenomena are interpreted holistically. Therefore, a case study was a suitable methodology for this study exploring the phenomenon of washback.

Participants

One hundred fifty high school students (75 males and 75 females) from two high schools in Saudi Arabia were invited to participate in this study. Purposive sampling was applied as it is a suitable method to this study so that only students majoring in science were included. This sampling approach was taken because the experiences of students majoring in science are different from other students, as they have taken both the GAT and the SAAT. Additionally, the students were selected from highly ranked schools in terms of students' scores on these tests.

All of the participants in this study had taken the SAAT and GAT at least once. The majority of the students (112) were in the last year of high school, whereas the rest (38) were in the second year. Consent forms were given to the students prior to the study. Data collection tools included semi-structured questionnaires and semi-structured interviews. Methodological triangulation was utilized to validate data, avoid bias, and obtain objectivity. Initial agreement was secured by the researcher from 150 students. Of the 150 students, 103 (56 females and 47 males) participants opted to complete the questionnaires (with 68% completion rate). Then, semi-structured interviews were conducted by the researcher with eight students (four males

and four females). The scores of these students were the highest among the whole sample and they took the SAAT and GAT only once during their studies.

Context of the Study

This study was conducted at two high schools in Al Hasa city in Saudi Arabia. The data collection process lasted for four months. The purpose of the construction of the semi-structured questionnaire was to explore the students' perceptions with respect to the SAAT and GAT positive washback. Likewise, the interviews were conducted to obtain an in-depth understanding of students' perceptions of the tests in order to verify the viewpoints that were mentioned in the questionnaires. Particular attention was paid to including relevant items through trials and expert reviews with a small sample of participants. All of the students completed the questionnaire through an online website, whereas the interviews were face-to-face.

Data analysis

The data were analyzed inductively in this qualitative study. The study followed the content analysis method by utilizing themes as the unit of analysis. In this study, the themes which emerged from the semi-structured interviews and the semi-structured questionnaires were coded by the researcher and analyzed through the NVivo program. Inferences from the data were summarized in order to find patterns and any existing relationships between data segments. Actually, it is beyond the scope of the study to build a new theory through a grounded theory approach. Therefore, content analysis was used to confirm a pre-existing theory. The incorporation of both methods (questionnaires and interviews) helped the researcher capture the wider picture of the research inquiry. Interview data were fully transcribed prior to the analysis stage. Then, for the purpose of validation, the interview scripts were given back to the interviewees.

Results

Questionnaire data

Students' scores

As mentioned above, 103 (56 females and 47 males) participants opted to complete the questionnaires. Their scores ranged from 68 to 94 on the GAT and from 57 to 89 on the SAAT. This variation in the mean scores

between these tests may show that the SAAT test was slightly more challenging to these students than the GAT.

Students' Repeat Performance

Most of the participants indicated that they had taken the tests more than once. The maximum number of attempts for the GAT was 5 times, while the maximum number of attempts for the SAAT was 2 times. Both of the tests' results are valid for five years. Students are allowed to take the GAT test when they are in their second year of secondary school, while the SAAT can be taken only in the third year or after graduation.

Positive Washback of the Tests

Almost all of the students who completed the questionnaires reported that the tests provided them with a unique chance to retake the them more than once; thus, the pressure was far lower than for the SSC tests. Specifically, they indicated that the retake procedure helped them to secure high scores because students' best scores on the tests are counted no matter how many times the tests are repeated. In addition, the students noted that the tests measured their achievement and readiness through general questions based on their secondary school courses (for the SAAT) as well as intermediate and elementary school courses (for the GAT). Importantly, the students noted that there is no pass or fail option. However, they also remarked that success on the tests required substantial focus on critical thinking and problem-solving strategies rather than just memorization.

Interpretation of Test Results

The majority of the students were satisfied with the way the results were interpreted and then used for admission purposes at universities. They reported that the SSC is no longer the sole indicator of academic performance or criterion for admission at universities. Instead, the SSC score now counts for 30%-60% of the total score needed for enrollment at universities. On the other hand, the GAT is given a weight of 30%-40%, whereas the SAAT is given a weight of 20%-50%. The weight of these tests varies from one university to another depending on the type of programs the university offers. Most students in this study reported that this change has facilitated the process of acceptance to universities. The students took the tests repeatedly until they felt they had performed well enough to be accepted to the universities they wanted to enroll in. Some students

continuously checked and compared the admission criteria for each university to be up to date with the latest requirements related to these tests.

Interview data

Eight students (four males and four females) were chosen from the entire group of 103 participants (47 males and 56 females) to participate in semi-structured interviews. The scores of these students were the highest among the whole sample and they took the SAAT and GAT only once during their studies. The intention was to choose students who had been positively influenced by the test.

Positive washback of the test

Soft skills

The majority of the students interviewed indicated that the GAT and SAAT gave them a chance to hone some of the soft skills that would be needed for future university study. One student said that she found the tests “helpful in terms of developing problem-solving...and like, you know, reasoning.” Similarly, another student added that “in such tests, especially the GAT you are required to have critical thinking to pass the test.” In addition, some students reported that they used other strategies to achieve excellence in these tests. One student noted that “time management is an important thing; otherwise, you will not succeed.” Likewise, there were other helpful tips that some students followed: “I learned how to use Arabic-Arabic dictionaries and something about answering easy questions first.”

These students experienced positive impacts from the tests on their performance through the use of such strategies. The tests seemingly require higher-order cognitive skills. The students also depended on themselves in solving complicated mathematical equations without using a calculator. They learned how to prioritize answering test items. The students' successful use of such skills may show that, at a larger level, the shift in education in Saudi Arabia from indoctrination to the development of soft skills has been successful in some ways. This shift may be consistent with global needs, as there has been a great emphasis placed “on soft skills over the last 5–10 years by both employers and graduates” (Succi & Canovi, 2019, p. 11).

Impact of the contextual variables on test scores

The majority of the students indicated in the interviews that the GAT and SAAT have decreased the influence of the SSC tests. Both the GAT and the SAAT are multiple choice tests, and some students reported

that this type of the question does not depend very much on memorization. For example, one student commented that “I don’t need to remember a long answer like the final exam questions in my school. Also, if I want to change my mind, it’s easy just select another option, and I can get a good grade.” This finding may corroborate Al-Hamly and Coombe’s (2005) suggestion that “effective use of answer changing as a test-taking strategy may result in students’ score improvement” (p. 520).

In addition, one student mentioned that she felt more secure with the GAT and SAAT than the SSC tests because “the machine is going to check our answers...and maybe you need to write a lot before, but now you choose the correct answer. That’s why I want to repeat the test more than one time.” She added, “I got a good grade because the answers are in the questions... Actually, I don’t take time to memorize things.” This method of testing seemed to help substantially in boosting the students’ scores on these tests, at least for the students who were interviewed. What these students said might be consistent with what Shohamy (2001, p. xi) found in her research during her early years in language testing, namely, that “the method of testing did, in fact, affect performance of test takers on tests.” This finding may suggest that there are some inherent features within a test along with other strategies students may tend to use that could lead to successful score boosting. Such practices may give rise to positive attitudes (i.e., the use of critical thinking and problem-solving strategies rather than just memorization) that the Ministry of Education may wish to promote in its students as form of educational reform. The aim is to foster independence in these students and to reduce the pressure on them to cram information. For this reason, it is believed that instruction at the university level should focus on “learning-how-to-learn rather than mastery of bodies of factual information” (Wong & Nunan, 2011, p. 144).

Additionally, these tests, according to the students, created positive washback by allowing students to retake the tests more than once. The result was that students felt that the pressure of taking the tests was lifted and that they had the opportunity to secure high scores, as their best scores would be counted. One student commented on this aspect, saying that “the best moment in my life is when... I think... I can again repeat the test.” He added, “I will sure try to do better next time...and get the best score.” Another student reported that she had tried “a lot of strategies to beat the tests.... I tried to answer the easy questions, and I looked for some words in the questions, but I did best when I repeated the test because I know which way was best to answer the questions.” It seems that when the students retook the tests, they improved their test-taking strategies as well as their

scores on the tests. Most students in this study increased their scores on both tests when they repeated the tests.

However, they found that score increases on the SAAT were far more difficult than those on the GAT. One of the reasons was that the maximum number of attempts to retake the SAAT among the students in the study was 2 times. Time was an issue because they could take the test before or after their final secondary school exams. Additional possible explanations for score gains were explained by some students as follows: "I think because the tests became easier to me when I repeated them" and "I went back and studied harder to improve my vocabulary." This finding seems to be consistent with Saida and Hattori's (2008, p. 190) suggestion in their study that the scores of the test are influenced by two factors "a difficulty factor for test items and an ability factor for examinees." In other words, score gains might be a result of a test being easier in one of the trials or of students' proficiency improving significantly. The students said that they had a list of the most frequent words that would appear in these tests, especially those words that were easy to remember. This student strategy is consistent with the findings of He and Godfroid (2019), who noted that "learners do not need to expend much mental energy to learn these words but can improve considerably in their communicative abilities" (p. 364).

Impact of the tests on the educational system

Seemingly, the GAT and SAAT tests were introduced in order to not only unify admission criteria among Saudi universities but also prompt educational reform. The aim was for school curricula to focus more on 21st century skills such as critical thinking and problem-solving strategies. The tests were one of the driving forces of change in the educational system. The majority of the students who took part in this study reported that school curricula have changed dramatically in the last two decades to include questions that, according to one student, required "deep thinking and, more importantly, reasoning." Specifically, she explained, "I was not familiar with these questions in elementary school...I think we should do now philosophical things, logic and think together." Such tasks were prevalent throughout all stages of education but differed in difficulty and focus. The development of skills through these tasks may have prepared students to take the GAT and SAAT tests.

Equally importantly, most students noted that the test questions depended on materials to which they had been exposed earlier in their studies. Therefore, the content of the tests was relevant and covered some of the topics they had already taken. However, these tests, according to one

student, “needed little preparation in advance, but you know, you cannot do this in one night like our courses in school because these tests are about all the things we had in these 12 years.” It seems that the introduction of such tests has raised students’ awareness in terms of early preparation and the subsequent benefits. Another student mentioned, “I think this is what I am supposed to do when I go to university...I mean, prepare early for my courses.”

However, one student said, “[Some students] don’t take it seriously...This is their last chance after school, I mean to prove they are good.” In other words, students who fail to obtain a high GPA in school can compensate for this loss through the GAT and SAAT tests because these tests are given a weight of 70% of the total score (30% for the GAT and 40% for the SAAT), whereas the SSC score is given only a 30% weight. Therefore, there is a high likelihood that students who have failed to achieve a high SSC score can achieve a high GAT or SAAT score. Actually, such a student performance trend seems to be consistent with the findings Wiggins (1998), who asserted that the aim of assessment should be to develop and improve student performance instead of just measure it.

Interpretation of test results

Fairness of the tests

Almost all of the students who participated in this study expressed their satisfaction with the way the test results are interpreted and then used for admission purposes at universities. Life chances for test takers have changed dramatically since the introduction of these tests. Students are now admitted to universities based on standardized measures of their academic performance. Two decades ago in Saudi Arabia, there was an overreliance on the SSC score as the sole indicator of performance. Teacher-made examinations (which lacked essential test qualities) may have helped some students with average performance outperform others and compete with excellent students because their grades were boosted so that they could secure admission into universities. However, when the GAT and SAAT were introduced, due to the poor outcomes of teacher-made examinations, average students started to suffer. One student said, “some students got high grades in secondary school, but they failed the Qiyas tests.” This finding is consistent with that of Siddiek (2011, p. 63), who noted that there were some “students who got more than 90% on their achievement tests but they got around 60% on the Qiyas test.”

Importantly, most of the students who participated in this study were happy with the current situation. They noted, "We feel we are equal, not because of my family, you give me high marks." The GAT and SAAT tests have reduced score inflation; such practices previously resulted in a deterioration in students' performance once they had been accepted into university, even though they had achieved high SSC scores. Siddiek (2011, p. 60) asserted that examinations play an important role in determining "the degree of individual fortune in getting well paid job according to his/her qualifications rather than his political, social, or ethnic affiliation." What the students described may also be consistent with Klasik and Strayhorn's (2018, p. 341) doubts about the readiness of US high school students' GPAs, as they "appear to paint too rosy a picture"; because "there appears to be grade inflation that varies by race/ethnicity that overestimates the college readiness of students."

Setting cut-off scores

According to the President of the Qiyas Center in Saudi Arabia, "the admission percentage in Saudi universities of secondary school graduates reached up to 90%" (Qiyas, 2013). Therefore, there was a pressing need to control this influx of students through standardized tests. The universities had the right to determine the best candidates who stood out compared to other students. Each university was given the freedom to set its cut-off scores depending on the type of program. There were already entrance exams for most Saudi universities, so the aim was to standardize the admission tests through valid measures of performance. Thus, the Qiyas standardized tests did not introduce something new, but rather unified the admission criteria across Saudi universities.

Most students in this study found that when the Qiyas Center eliminated the practice of setting cut-off scores for universities, the enrolment process was facilitated. One student noted that "this difference in scores interpretation has opened the doors for all students to take a place at a university and allowed the best students to be accepted into prestigious universities." This student may perhaps have witnessed the suffering of the disadvantaged students in the past who were refused entry based on unreliable criteria. Moreover, some students seem to have been preoccupied with the scores they had received on the tests. This preoccupation may have been because, as one student explained, "universities keep raising their standards," which could be why some students continued to retake the tests until they had the required composite score. This finding supports the findings of Ormond (2019), who noted in her study that educational systems

that want to produce globally competitive students need to “raise standards and monitor delivery of curriculum” (p. 150). Other students were also searching for universities that would allow them to enroll in similar programs with lower composite scores. Universities have now begun to base their decisions about the readiness of their students on the scores of these tests.

Conclusion

The current case study qualitatively explored the positive washback of testing in Saudi Arabia, focusing on secondary school students’ perceptions of national exams, specifically the GAT and SAAT. The findings of the study show that the GAT and SAAT have had a positive impact on secondary school students. This positive impact was observed in students’ low-pressure test-taking experiences, as they were allowed take the tests more than once, which ultimately provided them a unique chance to achieve high scores. In addition, the findings of this study suggested that the method of testing helped greatly boost the students’ scores on these tests, at least for these students. Above all, there was no pass or fail option on these tests. Nonetheless, success on the tests required substantial focus on critical thinking and problem-solving strategies rather than just memorization. The GAT and SAAT have positively impacted students’ performance in honing soft skills that are needed for future study at university. Importantly, most students felt that the pressure of enrollment in universities was lifted because the SSC is no longer the sole indicator of academic performance. In other words, students who failed to obtain a high GPA in school could compensate through the GAT and SAAT tests.

The findings of this study align with the presence of washback as an emerging theme in the field of assessment. However, the researcher urges caution when interpreting the results of this study because of the limited number of students who participated. In addition, the study was conducted in two schools in Saudi Arabia. Therefore, the results cannot be generalized to other contexts. Furthermore, the study utilized two data collection methods. One way for future research to mitigate this limitation is to have a larger sample of stakeholders from different contexts by using different data collection tools. Tzagari and Cheng (2017) noted that studies about washback of one test in a particular context by one researcher can still provide valuable insights about that context “however, it would be best if groups of researchers could work collaboratively and cooperatively to carry out a series of studies around the same test within the same or across

educational contexts” (p. 367). The results of this study have implications for GAT and SAAT stakeholders. For positive washback to occur, it is recommended that the GAT and SAAT continue to assess students' higher-order cognitive skills, encourage the fostering of independence, and help students prioritize answering test items. Test preparation and test retaking played key roles in promoting score gains and thus positive washback.

Acknowledgements

The author acknowledges the Deanship of Scientific Research at King Faisal University for the financial support under Nasher Track (Grant No. 186207).

References

- Al-Alwan, I. A. (2009). Association between scores in high school, aptitude and achievement exams and early performance in health science college. *Saudi Journal of Kidney Diseases and Transplantation*, 20(3), 448–453. Available from: <https://www.sjkdt.org/text.asp?2009/20/3/448/50777>
- Alanazi, N. K. (2014). *The validity of the general aptitude test (GAT) to predict male Saudi students' academic performance in first semester*. Unpublished PhD Dissertation, University of Northern Colorado. ProQuest Dissertations & Theses Global. <https://search.proquest.com/docview/1629828803?accountid=27804>
- Al-Hamly, M., & Coombe, C. (2005). To change or not to change: Investigating the value of MCQ answer changing for Gulf Arab students. *Language Testing*, 22(4), 509–531. <https://doi.org/10.1191/0265532205lt317oa>
- Al-Owidha, A. A. (2013). *The Saudi standardized achievement admission test (SAAT); A study of construct-related validity evidence* (TR038). Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.
- Al-Qataee, A. (2012a, March 21). *Test length and precision: How precise the GAT will be if the test is shortened?* (TR008). Riyadh, Saudi Arabia: National Center for Assessment in Higher Education. <https://etec.gov.sa/ar/Researchers/Research-Studies/Studies/1Full%20verses%20short%20form%20tests11%20final%20008.pdf>
- Al-Qataee, A. (2012b, August 3). *Differential validity of the GAT admission test in Saudi Arabia* (TR010). Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.
- Alazizi, E. (2016). The impact of a training program on the performance of a sample of students in the general abilities test in the province of Mahd Al-

- Thahab. *The Educational International Specialized, Jordan*, 5(7), 1–17.
<https://platform.almanhal.com/Files/2/100384>
- Aldahesh, A., Alnamer, M. & Alher, M. (2015). The impact of the students' GPA in high school and the score in the GAT and SAAT on the GPA at university. *Journal of Scientific Research in Education, Egypt*, 16(5), 477–494.
https://jsre.journals.ekb.eg/issue_2656_2709.html
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115–129. <https://doi.org/10.1093/applin/14.2.115>
- Alghamdi, R. (2017). The reasons of the gap between results of female students tests in high school and results of general aptitude and achievement tests from the perspective of students, teachers, managers, supervisors. *Journal of Culture and Development, Egypt*, 16(101), 73–117.
<http://mcd.arkanorg.com/?s=%D8%A7%D9%84%D9%81%D8%AC%D9%88%D8%A9>
- Alghamdi, S. (2012). The GAT and its relationship with future anxiety on a sample of secondary school students in Saudi Arabia. *Journal of College of Education, Egypt*, 23(90), 120–152. <https://content.mandumah.com/download?t=006748f28127725121603ef514e5211231080597&f=yIujVbp8RoaWsleq86KV1xryRgU1ti3hulA87odFJms=&s=1>
- Almehrej, H. (2009). The general aptitude test from the perspective of university students. *Journal of Arabic Studies*, 8(2), 205–262.
<http://search.mandumah.com/Record/82517>
- Alnahdi, G. H. (2015). Aptitude tests and successful college students: the predictive validity of the general aptitude test (GAT) in Saudi Arabia. *International Education Studies*, 8(4), 1–6. DOI:10.5539/ies.v8n4p1
- Alnowibet, K., & Ahmad, S. (2015). Saudi school assessment system for predicting admissions to science colleges. *The Online Journal of Quality in Higher Education*, 2(1), 37–41. <https://www.tojqih.net/journals/tojqih/articles/v02i01/v02i01-05.pdf>
- Al Ruwaili, S. K. (2013). *The Low Performance of High School Students in the General Aptitude Test in Al Jouf Region in Saudi Arabia and its Treatment: An Analytical Study*. Unpublished MA Thesis, Yarmouk University. <https://dirasat.ju.edu.jo/ES/Article/FullText/11746?volume=45&issue=4>
- Alshehri, M. (2007). *The Predictive Validity of General Aptitude Test (GAT) and Secondary School Test (SST) on Grade Point Average (GPA) at Saudi Universities*. Unpublished MA Thesis, Mutah University. <https://dirasat.ju.edu.jo/ES/Article/FullText/11746?volume=45&issue=4>
- Alshehri, M. (2010). The relationship between the first year mathematics at university and the GAT, SAAT, and mathematics of secondary school.

- Education Journal for Educational, Psychological and Social Research*, 144(3), 65–80.
<https://content-mandumah-com.sdl.idm.oclc.org/download?t=9aad857efb10f0cf2df92633c770075f101746bb&f=nPAYP9haj/jK3vOZPhUuds7AtXO0nG8NXoOVg3eKllw=&s=1>
- Alshemeri, M. M. (2015). *Equating Scores for Different Forms of Ability Test Among Secondary School Students in Saudi Arabia*. Unpublished MA Thesis, Yarmouk University. <https://search-mandumah-com.sdl.idm.oclc.org/Record/741294>
- Alshumrani, S. A. (2007). *Predictive validity of the general aptitude test and high school percentage for Saudi undergraduate students*. Unpublished PhD Dissertation, University of Kansas. ProQuest Dissertations & Theses Global. (304857763). Retrieved from https://search-proquest-com.sdl.idm.oclc.org/pqdtglobal/docview/3048577_63_/CE6F73F470B041CBPQ/8?accountid=142908
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Balkyoor, S. M. (2008). *The Efficacy of CTT and IRT Equating Methods to the Forms of the General Ability Test (GAT)*. Unpublished PhD Dissertation, Umm Al-Qura University.
- Bobsheet, A., & Al-Ghamdi, A. (2015). Perceptions of students at the university of Dammam towards general aptitude tests. *The Saudi Journal for Higher Education, Saudi Arabia*, 13, 137–168.
<http://chers.moe.gov.sa/ar/Publications/Publications1/HEJ-0013-Ar.pdf>
- Broadfoot, P. (1996). *Education, assessment and society: A Sociological Analysis*. Buckingham: Open University Press.
- Cheng, L., Sun, Y., & Ma, J. (2015). Review of washback research literature within Kane's argument-based validation framework. *Language Teaching*, 48(4), 436–470. DOI: <https://doi.org/10.1017/S0261444815000233>
- Cheng, L., Watanabe, Y., & Curtis, A. (2004). *Washback in language testing: Research contexts and methods*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Davies, A., Brown, A., Edler, C., Hill, K., Lumley, T., & McNamara, T. (1999). Dictionary of language testing. In *Studies in language testing 7*. Cambridge: Cambridge University Press.
- Deygers, B., Van den Branden, K., & Van Gorp, K. (2018). University entrance language tests: A matter of justice. *Language Testing*, 35(4), 449–476.
<https://doi.org/10.1177/0265532217706196>
- Dimitrov, D. M. (2013a, January 22). *Teacher tests-educational skills: Equating of test forms*. Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.

- Dimitrov, D. M. (2013b, March 12). *General Aptitude Test-verbal: Testing for dimensionality and validation of factorial structure* (TR011). Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.
- Dimitrov, D. M. (2013c, April 29). *General Aptitude Test -quantitative: Testing for dimensionality and validation of factorial structure* (TR012). Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.
- Dimitrov, D. M. (2013d, May 16). *General Aptitude Test-verbal: IRT and true-score analysis* (TR013). Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.
- Dimitrov, D. M. (2013e, June 30). *General Aptitude Test-quantitative: IRT and true-score analysis* (TR014). Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.
- Dimitrov, D. M. (2013f, August 19). *Latent class analysis of GAT-quantitative data* (TR015). Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.
- Dimitrov, D. M. (2013g, September 25). *Latent class analysis of GAT-verbal data* (TR016). Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.
- Dimitrov, D. M. (2013h, December 13). *Testing for unidimensionality of GAT data* (TR029). Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.
- Dimitrov, D. M. (2015). *Psychometric features of the Arabic and English versions of the General Aptitude Test (GAT): An initial investigation with large-scale assessment data in Saudi Arabia* (TR095). Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.
- Dimitrov, D. M., & Shamrani, A. R. (2015). Psychometric features of the General Aptitude Test–verbal part (GAT-V). *Measurement and Evaluation in Counseling and Development*, 48(2), 79–94. <https://doi.org/10.1177/0748175614563317>
- Dimitrov, D. M., Raykov, T., & Al-Qataee, A. A. (2015). Developing a measure of general academic ability. *Educational and Psychological Measurement*, 75(3), 475–490. <https://doi.org/10.1177/0013164414543179>
- Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. London: Falmer Press.
- Green, A. (2007). *IELTS washback in context: Preparation for academic writing in higher education*. Cambridge: Cambridge University Press.
- Hamdan, A., & Al-Hattami, A. (2016). The accuracy of predicting university students' academic success. *Journal of Saudi Educational and Psychological Association*, (54), 173–182. https://gesten.ksu.edu.sa/sites/gesten.ksu.edu.sa/files/8_0.pdf

- Hamp-Lyons, L. (1997). Washback, impact and validity: Ethical concerns. *Language Testing*, 14(3), 295–303. <https://doi.org/10.1177/026553229701400306>
- He, X., & Godfroid, A. (2019). Choosing words to teach: A novel method for vocabulary selection and its practical application. *TESOL Quarterly*, 53(2), 348–371. <https://doi.org/10.1002/tesq.483>
- Hughes, A. (2003). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Jehangir, K. (2015). *Investigating the impact of proctor background characteristics on achievement scores in GAT (TR087)*. Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.
- Jehangir, K. (2016a, April 3). *IRT model fit/residual analysis comparisons between 5 test forms of the computer based GAT quantitative test versus 5 corresponding test forms of the paper and pencil based GAT quantitative test (TR151)*. Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.
- Jehangir, K. (2016b, August 14). *IRT model fit/residual analysis comparisons between 5 test forms of the computer based GAT verbal test versus 5 corresponding test forms of the paper and pencil based GAT verbal test (TR154)*. Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.
- Khaniya, T. R. (1990). *Examinations as Instruments for Educational Change: Investigating the Washback Effects of the Nepalese English Examinations*. Unpublished PhD Thesis. Edinburgh, University of Edinburgh. Edinburgh Research Archive. <http://hdl.handle.net/1842/9862>
- Klasik, D., & Strayhorn, T. L. (2018). The complexity of college readiness: Differences by race and college selectivity. *Educational Researcher*, 47(6), 334–351. <https://doi.org/10.3102/0013189X18778598>
- Kunnan, A. J. (2004). Test fairness. In M. Milanovic, C. Weir, & S. Bolton (Eds.), *Europe language testing in a global context: Selected papers from the ALTE conference in Barcelona* (pp. 27–48). Cambridge, UK: Cambridge University Press.
- Lamprianou, I. (2013a, October 11). *Investigation for exposed items in the context of the General Aptitude Test (GAT) (TR023)*. Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.
- Lamprianou, I. (2013b, December 1). *Investigation for potentially exposed items in the context of the verbal section of the General Aptitude Test (GAT) (TR030)*. Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.
- Lamprianou, I. (2014a, September 5). *Confirmatory factor analysis as a method for assessing reliability and validity of the General Aptitude Test (GAT) for postgraduate students (TR050)*. Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.
- Lamprianou, I. (2014b, October 4). *The dimensionality of the investigation and general prosecution scale: The General Aptitude Test (TR057)*. Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.

- Lamprianou, I. (2014c, November 9). *Factorial invariance and latent mean differences of scores on GAT across school types* (TR062). Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.
- Lamprianou, I. (2014d, December 21). *Factorial invariance and latent mean differences of scores on GAT across gender* (TR063). Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.
- Lamprianou, I. (2015a, November 21). *Examining the effectiveness of the distractors in multiple-choice items using the differential distractor functioning (DDF) on SAAT scores across gender* (TR091). Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.
- Lamprianou, I. (2015b, December 17). *Factorial invariance and latent mean differences of scores on SAAT across gender* (TR092). Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.
- Lee, M., & Aluhiedan, S. (2013). *A comparison of students' performance in Standard Achievement Admission Test (SAAT) across educational administrations, types of school, genders and nationalities* (TR025). Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241–256. <https://doi.org/10.1177/026553229601300302>
- Ormond, B. M. (2019). The impact of standards-based assessment on knowledge for history education in New Zealand. *Assessment in Education: Principles, Policy & Practice*, 26(2), 143–165. <https://doi.org/10.1080/0969594X.2018.1432564>
- Qiyas (2013). Measurement can never be an obstacle to admitting students at universities [Video file]. Resource document. National Center for Assessment. Retrieved from <https://www.youtube.com/watch?v=qljmyqMZKH0>
- Saida, C., & Hattori, T. (2008). Post-hoc IRT equating of previously administered English tests for comparison of test scores. *Language Testing*, 25(2), 187–210. <https://doi.org/10.1177/0265532207086781>
- Saif, S. (2006). Aiming for positive washback: A case study of international teaching assistants. *Language Testing*, 23(1), 1–34. <https://doi.org/10.1191/0265532206lt322oa>
- 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. <https://doi.org/10.1080/15434301003664196>
- Shohamy, E. (2001). *The power of tests: A critical perspective of the uses of language tests*. Harlow: Longman.
- Siddiek, A. (2011). Standardization of the Saudi secondary school certificate examinations and their anticipated impact on foreign language education.

- International Journal of Humanities and Social Sciences*, 1(3), 57–64.
<http://www.ijhssnet.com/view.php?u=>
<https://www.ijhssnet.com/journals/Vol.1No.3March2011/9.pdf>
- Sideridis, G. D. (2014). *Assessing differential achievement scores on the GAT due to examination center characteristics* (TR048). Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.
- Sideridis, G. D. (2016). *Effects of testlet exposure on students' performance* (TR158). Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.
- Sideridis, G. D., & Tsousis, I. (2013a, December 15). *DIF analysis for item and test on the NCA tests: The General Ability Test (GAT) art major* (TR035). Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.
- Sideridis, G. D., & Tsousis, I. (2013b, December 30). *DIF analysis for item and test on the NCA tests: The General Ability Test (GAT) science major* (TR036). Riyadh, Saudi Arabia: National Center for Assessment in Higher Education.
- Sideridis, G. D., Tsousis, I., & Al-harbi, K. A. (2015). Multi-Population Invariance With Dichotomous Measures: Combining Multi-Group and MIMIC Methodologies in Evaluating the General Aptitude Test in the Arabic Language . *Journal of Psychoeducational Assessment*, 33(6), 568–584.
<https://doi.org/10.1177/0734282914567871>
- Sinclair, J., Fox, G., Moon, R., & Stock, P. (1987). *English language dictionary*. London, UK: Collins.
- Succi, C., & Canovi, M. (2019). Soft skills to enhance graduate employability: comparing students and employers' perceptions. *Studies in Higher Education*.
<https://doi.org/10.1080/03075079.2019.1585420>
- Sulphey, M. M., Al-Kahtani, N. S., & Syed, A. M. (2018). Relationship between admission grades and academic achievement. *Entrepreneurship and Sustainability Issues*, 5(3), 648–658. DOI 10.9770/jesi.2018.5.3(17)
- Taylor, L. (2005). Washback and impact. *ELT Journal*, 59(2), 154–155.
<https://doi.org/10.1093/eltj/ccj030>
- Tsagari, D., & Cheng, L. (2017). Washback, impact, and consequences revisited. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment* (pp. 359–372). Cham: Springer International Publishing.
- Wall, D. (1997). Impact and washback in language testing. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education 7: Language testing and assessment* (pp.291–302). Dordrecht: Kluwer Academic.
- Wiggins, G. (1998). *Educative assessment*. San Francisco, CA: Jossey-Bass Publishers.
- Wilkinson, A. (1968). The testing of oracy. In A. Davies (Ed.), *Language testing symposium: A psycholinguistic perspective* (pp. 117–132). Oxford, UK: Oxford University Press.

- Wiseman, S. (1961). *Examinations and English education*. Manchester: Manchester University Press.
- Wong, L. L. C., & Nunan, D. (2011). The learning styles and strategies of effective language learners. *System*, 39(2), 144–163. <https://doi.org/10.1016/j.system.2011.05.004>