

Research Article

The Influence of High-Stakes English Examinations on Students' Out-of-Classroom English Learning Practices: A Comparative Study

Getachew Desalegn ¹, Rufael Disassa,¹ and Tamene Kitila²

¹Department of English Language and Literature, Institute of Languages Study and Journalism, Wollega University, Nekemte, Ethiopia

²Department of English Language and Literature, College of Humanities, Language Studies, Journalism and Communication, Addis Ababa University, Addis Ababa, Ethiopia

Correspondence should be addressed to Getachew Desalegn; getachewd@wollegauniversity.edu.et

Received 27 November 2022; Revised 6 October 2023; Accepted 7 October 2023; Published 30 October 2023

Academic Editor: Shi Yin

Copyright © 2023 Getachew Desalegn et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Exams with high stakes may affect test takers' learning endeavors. Given the students' different academic backgrounds, it is not yet clear how much of an effect the tests will have on their out-of-classroom learning practices. This study, thus, aimed to close the knowledge gap by employing a mixed method of embedded research design and collecting data via questionnaire, focused group discussion (FGD), and document analysis. The questionnaire was administered to a stratified random sample of 94 12th grade students. The students with their one-semester academic achievements were selected from two secondary schools found in East Wollega zone, Oromia Regional State, Ethiopia. Data analysis involved descriptive statistics, multivariate and one-way ANOVA. The data gathered through document analysis and FGD were to substantiate the questionnaire. The contents of the three consecutive years of past Ethiopian Secondary School Leaving Certificate English Examination (ESSLCEE) questions were analyzed quantitatively. The qualitative method was used to conduct FGD with the selected participants from each school. The recorded data were subsequently transcribed, translated, analyzed, and discussed thematically. The study found significant differences between students of low-achieving groups ("Fair" and "Satisfactory" scorers) and high-achieving groups ("Very Good" and "Excellent" scorers) in studying non-ESSLCEE-related learning activities out-of-classroom. However, no statistically significant differences were observed between low-achieving groups and high-achieving groups regarding practicing ESSLCEE-related language components. The washback of high-stakes ESSLCEE on students' out-of-classroom English learning practices was observed regardless of the students' specific academic achievement groups they belonged to. The results suggest that there are differences in the impact of high-stakes exams on the efforts that students from different academic backgrounds make for out-of-classroom learning practices.

1. Introduction

Attempts to respond to the paucity of washback studies on learners and learning have been on the rise [1–8]. A study by Allen [1] investigated washback of the IELTS test on learners in the Japanese tertiary context is one example of such studies. The finding shows that the IELTS test created positive washback on learners' language ability and test preparation strategies. Similarly, Dong and Liu [2] studied the impact of learners' perceptions of a high-stakes test on their learning motivation and learning time allotment. The authors reported that students' positive test perceptions predicted

their intrinsic motivations better than did their negative test perceptions. The emerging studies are, however, biased toward classroom contexts [3, 5, 9], placing less emphasis on the out-of-classroom learning. That is not much seems to have been made to avoid a notion that washback on learning is rooted only in classrooms. To date, only a few researchers have attempted to extend washback studies beyond the classroom territory [7, 10–12].

Teaching, learning, and testing are the three essential and inseparable classroom practices that exist in education [13–15]. Testing is significant in measuring learners' language ability. It serves as a criterion to determine students'

fate for further education. It also evaluates the quality of education and controls nepotism [16–18]. These days, the consequential effect of testing on teaching and learning known as washback [19–24] is the concern of many educational researchers [25, 26]. This concern has been documented in literature, for example, [27–30]. Stoneman [30], for example, argued that the tests in reality exert greater power and have impact on its stakeholders across many walks of their lives. Mahmud [27] also stated that tests are powerful determiners of what happens in the classroom in shaping and influencing the teaching and learning process. The washback of testing especially the high-stakes [29, 31] on teaching–learning has become the ever-existing phenomenon [27, 29, 31, 32] with its multidimensional and complex nature [27, 33, 34].

In language testing, a large body of research has increasingly been undertaken on the washback of high-stakes tests on teaching–learning and the stakeholders, for example, [35–37]. However, not much work seems to have been done to show the effect of high-stakes tests on learners compared to the studies carried out on teaching [30, 38, 39]. It should be noted that learners’ lives are mainly and directly influenced by the washback of tests [18, 30, 40, 41].

Tsang [7] examined the mediating factors that affect students’ learning beyond the classroom in Hong Kong. In Turkey, Buyukkeles’ [10] study revealed that tests had no significant washback on students’ intrinsically motivated learning behaviors regardless of their language proficiency. Another study conducted by Pan [11] in Taiwan explored learners’ washback variability in standardized exit tests, and reported that high-proficiency students may engage in more learning activities and have more positive views of examinations than low-proficiency students. Yet another study in China by Zhan and Andrews [12] found that students’ perceptions of test importance affected their learning time allocation.

In the context of the current study, no study seems to have been carried out to tell us the extent of the influence of high-stakes Ethiopian Secondary School Leaving Certificate English Examination (ESSLCEE) on the students’ out-of-classroom English-learning practices [42–44]. Examining the variations of washback of high-stakes ESSLCEE among students of different academic achievement levels provides an important opportunity for English as a foreign language (EFL) teachers and other education stakeholders to gain deeper understanding of the link between the washback of tests on learning endeavors the students make beyond the classroom context. This realization led to the requirement for the current investigation.

2. Review of Literature

2.1. Testing and Its Washback on Language Teaching–Learning. Testing is indispensable in the context of language education. It is central to language teaching and learning [14, 15, 18, 45]. According to McNamara [45], language tests share a significant role in many people’s lives, acting as gateways at important transitional moments in education, in employment, and in moving from one country to another. Other scholars also voiced the contribution of language testing for testing language

learners’ ability, working as a criterion for admission of students to higher education, evaluating quality of education, and controlling nepotism in the allocation of scarce opportunities [16, 17, 46]. Thus, teaching, learning, and testing co-exist in the world of education [13–15]. Heaton [13] noted that the three constituents are so closely interrelated such that it is practically impossible to work in either field without being constantly concerned with the other. In the togetherness of the three entities, the leading power of testing over teaching and learning was gradually noticed with different terminologies, but the same concept; primarily, with case of “validity” [47], then with the concern of “measurement-driven instruction” [48], and with the issue of “curriculum alignment” [49]. Later, and now, it has been famed as “backwash” or “washback” with the consequential effect of testing on learning and teaching specific to classroom [19] and called “test impact” with its inclusiveness of the context and stakeholders beyond the classroom [49]. Following Alderson and Wall’s [19] call for empirical washback studies researching into the effects testing has on teaching and learning, various models and concepts [6, 17, 20, 50–53] have been developed to better understand the complex nature of washback [23, 33, 34, 41, 50]. To discuss each of these theoretical studies, it will be beyond the scope of this study. To the purpose of this work only, as the study aimed to investigate the washback of high-stakes ESSLCEE on learning, Alderson and Wall’s [19] washback hypotheses regarding what washback effects might look like on learners are stated below:

- (1) A test will influence learning;
- (2) A test will influence what learners learn;
- (3) A test will influence how learners learn;
- (4) A test will influence the rate and sequence of learning;
- (5) A test will influence the degree and depth of learning;
- (6) A test will influence attitudes to the content, method, and so forth, of learning;
- (7) Tests will have washback effects for some learners but not for others.

The above listed hypotheses highlight areas of learning that are generally affected by washback. The argument of the two scholars centered on the need to define various dependent variables in washback research to see their relationships with learning. In addition, the extent of the washback effect of tests may not be similarly practical on all learners. However, limited explanations about the rationale for possible variations of washback effects among different learners appear to be available in the literature [10, 11]. In connection with this, McNamara [45] pointed out that none of the 15 washback hypotheses mentioned factors that tell us how and why teachers and learners behave in certain ways in the classroom.

Empirical studies [34, 54] are entrusted with the responsibility of determining how tests affect teacher behavior, classroom practices, and test takers. Alderson and Hamp-Lyons [34] investigated the washback of test of English as a foreign language (TOEFL) on teaching, and it was reported

that TOEFL exerts an undesirable influence on language teaching. It was suggested that the need for more complex hypotheses about washback. Wall [54] studied the impact of high-stakes examinations on classroom teaching. The paper summarized what language testers have learned about test impact and discussed how tests interact with other factors in the testing situation. The reviewed literature highlights the multifaceted nature of test washback, which serves as an input for the current study, which examines the variability of test washback on students' English-learning practices beyond the classroom.

2.2. Washback of High-Stakes Tests on Learning. A large body of literature shows that the little emphasis had for long been placed on learners in the studies conducted on washback of high-stakes tests on teaching and learning [30, 38, 39, 55]. Even in the studies where the researchers tended to include learners, the involvement of learners was initiated by the need for providing complementary perspectives to the research conducted on teaching for data triangulation [34]. This negligence has been observed despite the fact that tests have critical effects on learners' lives. More recently, however, washback researchers have begun recommending the need to explore the impact of tests on learners [38, 39, 53] because the subjects are directly affected by the tests.

Apparently, in response to the washback researchers' call, washback studies that focused on different perspectives of learning have begun to emerge. Some such studies have focused on students' views of tests, for example [2, 24, 35, 56–60], others predominantly worked on the learning practices of students [1, 3, 5–8, 11, 61–64], whereas still other studies tended to give attention to both learning viewpoints and practices of test takers [4, 10, 11, 65–67]. Some of the studies referred appear to have reported contradictory results. For instance, Pan [11] investigated that intermediate and high-proficiency students spent much time on both test preparation and language-skill building activities than the low-proficiency students. Contrary to Pan's finding, Buyukkeles [10] reported as the lower proficient students more frequently involved in certain nontest-related activities. This is an indication of the need for yet more work in the context of washback studies that involve learners.

Some washback studies reported that tests influence what is learned, but there is no information in their report on how what is learned can be influenced by tests. Other studies revealed that the techniques students used to prepare for different exams were similar [30]. They reported that reading textbooks, memorizing vocabulary and idioms, going through previous exams, or relying on test prep books characterized their research subjects.

Tests have a variety of effects on students. The effects may vary according to the test takers' view of the tests or their different levels of language proficiency. For example, Stoneman [30], Shohamy et al. [24], and Tsagari [67] reported that learners' perceptions of the stakes and their perceptions of the status of the tests influenced the strength of the test effects. This means that (a) a high-stakes, high-status test promoted learning; (b) students spent more time engaging themselves in learning language skills that were

covered on the test than they did on lower-stakes or lower-status tests. Similarly, Tsai and Tsou [59] found that negative student opinions on the adoption of standardized exit tests led to a decrease in motivation to learn English because their classes were test-oriented, only enhancing their test-taking skills instead of their communicative competence.

The findings of washback studies that involved language proficiency appear to be mixed although they tend to contend that students' levels of English proficiency carried some weight in determining the extent of the effort, they were likely to make toward a test. According to Stoneman [30], Watanabe [60], and Chu [56], low-achieving students tended to be more worried about the test or test requirement than high-achieving students, and low-achieving students did not prepare for the test until the last minute or did not prepare at all.

As contended by Watanabe, a test of appropriate difficulty for the learner can positively affect their motivation to prepare for the test. In contrast, Ferman [68] and Shohamy et al. [24] found that students with lower abilities, given their belief that studying improves their scores, engaged in more intense learning for the test than did their counterparts. The higher-ability students, according to their report, were already eager to learn, even without the push of the test.

In particular, as available literature on washback studies in Ethiopia indicate, among few washback studies conducted in the country [42–44, 69–71], only three of them [42–44] concerned learning and the learners. Ayele [42] investigated that the wash-back effects of national higher education entrance examinations on students' English-language learning. He recommended that students be aware of the English course contents important for them to succeed in all other academic courses in their future career. Gashaye [43] conducted a study on the washback of grade 10 Ethiopian national English examination on students' practice. The result revealed that students were found practicing mainly grammar and technical aspects of writing and speaking disregarding using textbook due to the mediating factors such as students' ambition for success in the exam, awareness of the exam and teachers' exam-oriented teaching. Reta [44] explored the impact of Ethiopian high-stakes EFL tests on the role of teachers, learners, and parents. The findings of the study showed that the nature and content of the high-stake tests profoundly influenced the instruction practice of teachers, the learning practice of learners and the role of parents to the effect of focusing on grammar, vocabulary and reading comprehension at the expense of productive skills.

Significant contributions have been made by local and worldwide studies on the washback of examinations on learning and learners in reaction to the accusations of neglect made by the earlier washback researchers in the field. However, none of the research mentioned here examine the extent to which the out-of-classroom English study that students with various academic backgrounds engaged in relation to the results of English examinations. None of them made an effort to demonstrate how the students' academic achievement was related to their out-of-classroom English-learning practices. This research therefore aims to explore

the washback of high-stakes ESSLCEE variations that may exist among students on their learning practices out-of-classroom because of their academic achievement levels. Thus, to fill the research gap, the present study has been designed to seek answers to the following basic questions:

- (1) To what extent does high-stake ESSLCEE influence students' autonomous out-of-classroom English-learning practices?
- (2) Do students of different academic achievement levels report significant differences in practicing non-ESSLCEE-related language components?
- (3) Do students of different academic achievement levels report significant differences in practicing ESSLCEE-related language components?

3. Materials and Methods

3.1. Research Design. This study employed a mixed method of embedded research design. Researchers used mixed methods to blend data from both quantitative and qualitative sources and to better understand their study problem [72]. Mixed approaches combine the benefits of qualitative and quantitative data to validate quantitative results with qualitative data [73]. Combining and syncing many data sources might be useful for analyzing complex circumstances [74].

Both quantitative and qualitative data can be collected simultaneously with the use of an embedded research design [72]. The researchers' primary justification for choosing this design is that they wish to concentrate on the quantitative data regarding the impact of high-stakes ESSLCEE on students of various academic achievement levels while they learn English outside of the context of the classroom, and support it with the qualitative data.

To sufficiently investigate the quantitative results, two focus group discussions (FGDs) with six participants each were conducted after the original data—quantitative findings from a survey of 94 grade 12 students—[73]. When comparing various achieving groups, the magnitude of the impact of high-stakes ESSLCEE on students' out-of-classroom learning habits was described using a descriptive survey design, which offers a quantitative assessment of a population's patterns by evaluating a sample of that population [75]. Researchers used qualitative data analysis to support their quantitative data analysis findings [76]. To discover more about the effects of high-stakes ESSLCEE on students of various academic levels and how this affected their out-of-classroom English-learning habits, the researchers used appropriate FGDs.

3.2. Context and Participants

3.2.1. Context. In Ethiopia, English is taught as a compulsory subject from grade 1 to 8 (primary level), and serves as a medium of instruction starting from grade 9 [77]. The country has a fairly long history of conducting national examinations. The first national Examination of the country appeared in 1946 for grade 6 students under the name *London General Certificate of Examination* (GCE) [69]. Four

years later, in 1950, the *Ethiopian School Leaving Certificate Examination* (ESLCE) for grade 12 students was begun to be prepared as an experimental form [70] in parallel with London GCE. Starting from 1962, the Ethiopian Ministry of Education began to prepare the national examination independently [70, 71]. A new national examination scheme under the name of the *Ethiopian General Secondary Education Certificate Examination* (EGSECE) for grade 10 students was introduced in 2001, and then *Ethiopian Higher Education Entrance Examination* (EHEEE) for grade 12 students replaced the ESLCE starting from 2003. In 2019, the EGSECE in its turn was scraped without a replacement from the education system. The current high-stakes test of the country, ESSLCEE, has been serving to screen students for university education.

3.2.2. Participants. The study settings were two secondary schools in East Wollega zone, Oromia, Ethiopia. The zonal town of East Wollega, Nekemte, is located 320 km away from Addis Ababa. The specific names of the two schools considered in the study are Leka–Nekemte and Arjo secondary schools. The former is located in Nekemte town whereas the latter's distance from the town is 42 km. The two schools were first visited for permission to conduct the study in the schools. After getting permission from the school authorities, an appointment was made to sample the participants for questionnaire and FGD. The researchers received an ethics permission letter from Wollega University's College of Languages Study and Journalism's Research and Technology Transfer Post Graduate Office, with the following code numbers: ILSJ/98/2013.

3.3. Sampling. "Fair," "Satisfactory," "Very Good," and "Excellent" were used to categorize the 374 grade 12 students' levels of achievement. The category was using the standards set forth by the Ethiopian Ministry of Education (90–100 = "Excellent," 80–90 = "Very Good," 60–79 = "Satisfactory," 50–59 = "Fair," and 50 = "Poor"). The "Poor" achievement group is not included in the current study because no students in that category received scores below 50. There were 186 students from Arjo Secondary School and 188 students from Leka–Nekemte. Ninety-four individuals were chosen by stratified random sampling, with 47 coming from each school. At the time of the data collection, which took place during the academic year 2021–2022, the participants were grade 12 students. Each student who took part in the study was at least 18 years old. The details of the sampling are presented in Table 1.

3.4. Instruments. Three data gathering instruments, namely, questionnaire, FGD, and content analysis were used in the study. The questionnaire was the main data gathering instrument whereas the data from the other instruments: FGD and content analysis were employed for triangulation purpose.

A stratified sample of 94 students from the two secondary schools participated in the study was given the 11 closed-ended questionnaire items. The questionnaire was filled out by every participant. The very purpose of questionnaire is that it allows for extensive coverage with the least amount of time and money invested. It also allows for better geographic

TABLE 1: Summary of sampled students for questionnaire.

Schools' name	Total		Achievement categories							
			Fair (50–59)		Satisfactory (60–79)		Very good (80–89)		Excellent (90–100)	
	Pop	Sam	Pop	Sam	Pop	Sam	Pop	Sam	Pop	Sam
Arjo secondary school	186	47	51	13	87	22	36	9	12	3
Leka–Nekemte secondary school	188	47	48	12	84	21	40	10	16	4
Total	374	94	99	25	171	43	76	19	28	7

TABLE 2: Cronbach's α test result.

Dependent variables	Reliability from pilot test participants	Reliability from main study participants	No of items
EFL learning practices out-of-classroom	0.738	0.812	11

coverage and increases the validity of the findings by encouraging the selection of a sizable and representative sample [78]. Particularly in the context of the current study, questionnaire is helpful for assessing students' English-learning practices out-of-classroom. The results of studies by Mickan and Motteram [5], Buyukkeles [10], Pan [11], Zhan and Andrews [12], and a review of the washback literature were used to develop a questionnaire. In the survey, closed-ended questions predominated. The instrument, which consists of 11 total questions, was created to answer the central question about the effect of high-stakes ESSLCEE on students' out-of-classroom English-learning strategies. The five rating ranges for the products range from 1 for never to 5 for daily.

FGDs were also employed to investigate the students' out-of-classroom English-learning practices by triangulating the reported data with the data from the questionnaire. This facilitates the collection of comprehensive data regarding the impact of high-stakes ESSLCEE on students' out-of-classroom English-learning habits. Focus groups are especially useful for the current study because discussants are similar and friendly with one another, and discussion between discussants is likely to produce the most useful information [79]. It offers researchers a defined collection of facts and can deliver trustworthy, comparable qualitative data. Questions which regarded the influence of high-stakes ESSLCEE on students' out-of-classroom English-learning practices were included. Two groups, each with six participants drawn systematically from a range of academic achievement levels, participated in a FGD on two major areas of focus. The FGD conducted at Leka–Nekemte Secondary School took 39 min, and the one conducted at Arjo Secondary School took 30 min. The discussions were made in Afan Oromo, the language preferred by the participants.

From the schools, a sample of the previous ESSLCEE questions from 3 years in a row was collected. The systematic evaluation of recent documents was employed as data gathering tool [72]. The document data included the printed versions of the ESSLCEE questions from the previous 3 years in a row. It provides a useful function in the current study by supplementing the information gleaned from the questionnaire and FGD.

3.5. Validity and Reliability. The reliability and validity of the items are required to be verified prior to the actual data collection. The feedback from coworkers and the research supervisors was used to validate the questionnaire and FGD items. After the questionnaire and FGD items were translated into Afan Oromo, the researchers asked language specialists from the Department of Afan Oromo and Literature at Wollega University for feedback. Revisions were then made in response to the comments obtained.

Pilot testing was done using data from 45 grade 12 students who were enrolled in classes at the Burka-Jato, Dalo, and Kiba-Wacha secondary schools in Nekemte town to determine the validity of the questionnaire items in the context area. The analysis of the pilot research results was then presented at a seminar hosted by Wollega University's School of Post Graduate Studies. Before gathering data for the main study, the items were revised in light of feedback from participants and supervisors. The reliability test findings from the pilot and main study were examined using Cronbach α , and they were judged to be satisfactory [80] as shown in Table 2.

3.6. Research Procedure. The researchers received recommendation letters from Wollega University Post Graduate Office to gather data from the proposed schools. Following receipt of the letters, the letters were sent to the schools requesting their involvement in the study. Then, document was requested from the school record offices to register the lists of participant students those attending classes through the course of a semester. A questionnaire was given out after deciding which students would take part. Next, participants were chosen for the FGD from a variety of academic achievement categories, and the conversations were audio recorded. At the end, samples of three consecutive years of past ESSLCEE questions were collected.

3.7. Method of Data Analysis. In an embedded research design, the qualitative data were collected independently and analyzed to support the larger design (the quantitative data) [72]. To support the statistical findings, the researchers initially gave the quantitative statistical data before going over the main themes of the qualitative findings. The

TABLE 3: Descriptive statistics on the influence of ESSLCEE on students' learning practices.

<i>Non-ESSLCEE-related items</i>	<i>M</i>	<i>SD</i>	<i>ESSLCEE-related items</i>	<i>M</i>	<i>SD</i>
Reading from textbooks and authentic materials	2.10	0.843	Practicing ESSLCEE-related reading comprehensions	3.88	0.565
Studying English vocabularies	3.22	0.792	Practicing ESSLCEE-related vocabulary questions	4.18	0.586
Studying grammar and practicing its activities	3.34	0.498	Practicing ESSLCEE-related grammar questions	4.49	0.600
Practicing to improve oral conversations	2.74	0.655	Practicing ESSLCEE-related dialog questions	4.27	0.642
Practicing to improve writing skills	2.41	0.782	Practicing ESSLCEE-related writing questions	4.29	0.666
Watching and listening to English broadcasts	2.15	0.892			

validation of the questionnaire and coding served as the foundation for the data analysis procedure. Each questionnaire was graded and classified using the respondents' classifications of academic achievement: "Fair," "Satisfactory," "Very Good," and "Excellent." The data from the questionnaire were then encoded, tabulated, and its mean was calculated depending on the order of study questions using IBM SPSS statistics software, which has 25 versions. For the purpose of analysis, survey data were divided into two thematic categories: non-ESSLCEE-related and ESSLCEE-related items. Using descriptive statistics, the first research question was addressed. Multivariate, one-way ANOVA and post hoc test were used to analyze the data to answer the second and third research questions. All audio-recorded FGD courses were converted into text, coded to minimize their length, arranged according to themes that emerged from the collected empirical data. The analyses were made based on the two main themes emerged among the two groups. The first theme was with the discussions made regarding non-ESSLCEE-related out-of-classroom learning practices, and the second theme was with the discussions made on ESSLCEE-related out-of-classroom learning practices. The data from the samples of past ESSLCEE questions were analyzed using content analysis. The proportion of language components appear in the samples of these past ESSLCEE questions was tallied and described. The order in which the study questions were posed served as a guide for the data analysis.

4. Results

4.1. Influence of High-Stakes ESSLCEE on Students' Learning Practices Out-of-Classroom. To know about the extent to which the washback of high-stakes ESSLCEE affect students' English-learning practices out-of-classroom, quantitative data collected through questionnaire were analyzed using descriptive statistics. The descriptions were made comparing non-ESSLCEE-related items' mean scores' reports against with ESSLCEE-related items (Table 3). The questionnaire data were triangulated with FGD and documents to ensure its validity and reliability.

The left column of Table 3 first row has six consecutive items that are language development learning activities unrelated to the ESSLCEE. The next five consecutive items in the

same table's right column correspond to the language development learning activities associated to the ESSLCEE. The findings show that students replied more favorably to ESSLCEE-related questions than to non-ESSLCEE-related language components. For instance, it has been demonstrated that when students practiced reading comprehension questions connected to the ESSLCEE, they achieved a mean score of $M = 3.88$, $SD = 0.565$, as opposed to $M = 2.10$, $SD = 0.843$ when they studied reading comprehension materials unrelated to the ESSLCEE. Similarly, the mean value of $M = 2.41$; $SD = 0.782$ was recorded in response to the non-ESSLCEE-related writing skills activities, which is a low score when compared to the students' answers to the ESSLCEE-related writing skills questions ($M = 4.29$; $SD = 0.666$).

According to the statistics, responses to language development learning activities associated to the ESSLCEE had mean scores that were higher than those for language development learning activities unrelated to the ESSLCEE. The study found that students spent more study time on those ESSLCEE-related activities than on activities that were not related to the ESSLCEE. The outcome demonstrates that the impact of high-stakes ESSLCEE on students' outside-of-classroom English study time is noted.

Another important issue which is noted in the report is that the similarities and at the same time differences of the mean score in their rank order was found. For instance, the mean score recorded in responses to grammar activities of both non-ESSLCEE-related ($M = 3.34$; $SD = 0.498$) and ESSLCEE-related ($M = 4.49$; $SD = 0.600$) items showed as the mean scores registered put both the items inside the first rank order (Table 4). Conversely, the mean score recorded placed both of the reading comprehension exercises for non-ESSLCEE ($M = 2.10$; $SD = 0.843$) and ESSLCEE-related ($M = 3.88$; $SD = 0.565$) items in the last rank order (Table 4).

The findings indicate that students spent more time studying grammar exercises in both categories of language components unrelated to the ESSLCEE and those related to it, but they spent less time practicing reading comprehension in both categories. Because of the frequency with which language-related components featured in previous ESSLCEE questions, it is probable that students were more attracted to study grammar, vocabulary, speaking, and writing elements of activities.

TABLE 4: Rank order of language components within their mean score.

<i>Non-ESSLCEE-related items</i>	<i>M</i>	<i>Rank order</i>	<i>ESSLCEE-related items</i>	<i>M</i>	<i>Rank order</i>
Studying grammar and practicing its activities	3.34	1st	Practicing ESSLCEE-related grammar questions	4.49	1st
Studying English vocabularies	3.22	2nd	Practicing ESSLCEE-related writing questions	4.29	2nd
Practicing to improve oral conversations	2.74	3rd	Practicing ESSLCEE-related dialog questions	4.27	3rd
Practicing to improve writing skills	2.41	4th	Practicing ESSLCEE-related vocabulary questions	4.18	4th
Watching and listening to English broadcasts	2.15	5th	Practicing ESSLCEE-related reading comprehensions	3.88	5th
Reading from textbooks and authentic materials	2.10	6th			

TABLE 5: Composition of language components in ESSLCEE questions (2018–2020).

Year of examination	Reading comprehension	Vocabulary	Grammar	Dialogs	Writing	Total
Year 2020	22	16	38	26	18	120
Year 2019	14	16	26	34	30	120
Year 2018	14	17	26	32	31	120
Per (%)	13.88	13.61	25	25.55	21.94	100

The frequency with which language components appear in the content of previous ESSLCEE questions was examined for probative value (Table 5). For grade 12 students who are leaving high school and getting ready to enter higher education, 120 English items are prepared every year under the direction of the Ethiopian Educational Assessment and Examination Service. Reading, vocabulary, grammar, dialogs, and writing are among the five types of language components that are included in the items.

As shown, the 360 items from the three subsequent years (2018–2020) are distributed both numerically and proportionally. Reading comprehension questions account for 50 (13.88%), vocabulary questions for 49 (13.61%), grammar questions for 90 (25%), dialog questions for 92 (25.55), and writing questions for 79 (21.94) of the total ESSLCEE questions. Comparing the proportion of the ESSLCEE components, it is reported that the number of dialog, grammar, and writing skill assessment questions was higher in proportion than the rests. In the sample of ESSLCEE questions from the previous 3 years, reading comprehension and vocabulary make up the two language components with the lowest percentage among the five (reading comprehension = 13.88% and vocabulary = 13.61%). Reading comprehension and vocabulary make up the two language components with the lowest percentages among the five in the sample of ESSLCEE questions from the preceding 3 years (reading comprehension = 13.88% and vocabulary = 13.61%). The past researches done by Sato, [81], Xie [63], Zhan and Andrews [12], and Zhan and Wan [64] support the association of reports regarding reading comprehension it comprises the lowest in percentage in the past ESSLCEE and the mean score reported as the least in rank among other language components. The vocabulary report,

however, demonstrates that there is no connection between the proportional content of past ESSLCEE questions and the mean score reported on ESSLCEE-related activities. The reason is that students' participation in ESSLCEE activities involving vocabulary is rated second within their mean score, but having the lowest percentage of all language elements found in previous ESSLCEE questions. The findings made by Pan and Newfields [28] and Pan [82] are connected to this one.

The report from FGDs corroborated with the data obtained from quantitative data. Out of the 12 participants of the group discussions, majority of them had favored practicing specific English-language components related to ESSLCEE out-of-their classroom. The discussants shared their exposures in preparing for the coming high-stakes ESSLCEE as follows:

...I am attending tutorial classes which are arranged by some teachers. We pay some fee and our teachers teach us focusing on questions which may appear on the national examination (Student Participant 2).

I want to budget my after classroom study time for English subject by focusing on those appear on ESSLCEE. Still I am doing well on the grammar, writing and dialogue questions referring commercial books. However, it is difficult to get past ESSLCEE-related reading comprehension questions to practice; my school has no sufficient copies of the past exam papers; commercial books do not include the reading texts. Even to read from online, the internet access is very limited in our area (Student Participant 3).

TABLE 6: Comparative analysis of non-ESSLCEE-related items across the groups.

Items	Fair (50–59)	Satisfactory (60–79)	Very good (80–89)	Excellent (90–100)	Total
Reading from textbooks and authentic materials	1.72	1.65	3.05	3.57	2.1
Studying English vocabularies	2.92	2.81	4.16	4.29	3.22
Studying grammar and practicing its activities	3.16	3.07	4.05	3.71	3.34
Practicing to improve oral conversations	2.48	2.53	3.32	3.43	2.74
Practicing to improve writing skills	1.96	2.07	3.37	3.57	2.41
Watching and listening to English broadcasts	1.64	1.74	3.26	3.43	2.15
Total	2.31	2.31	3.54	3.67	2.66

I have no experience of budgeting for my study time concerning non-exam related English learning activities (Student Participant 5).

Personally, I read different books like “Extreme Series English” which help me to prepare for entrance exam... (Student Participant 7).

Out of my classroom, I practice doing questions from worksheets which contain compile of past ESSLCEE questions (Student Participant 8).

After the school, I study for the coming national examination relating the past ESSLCEE questions with the contents of the English textbooks (Student Participant 10).

I read “exams book” which my school awarded me (Student Participant 12).

On the other hand, some participants had experience in practicing questions unrelated to the ESSLCEE to enhance their general English proficiency. Following is how they described their exposures after class:

To improve my English language speaking ability, I read additional materials like “Gadaa Conversation”. I also read other books that focus on grammar to improve my knowledge in grammar (Student Participant 1).

I study English not only to pass the coming national examination, but also I work on the language to succeed in all academic courses I will take using English as a medium of instruction (Student Participant 6).

To improve my English, I practice speaking the language at home with some of my parents who are good at English. In addition, I use the internet and watch television programs in English (Student Participant 9).

To improve my English ability, I use different mechanisms like watching films in English and TV channels like BBC and Aljazeera (Student Participant 11).

Comparing the two themes that emerged from the participants’ comments on non-ESSLCEE-related and ESSLCEE-related topics shows that fewer number of participants claimed that they were actively working to improve their overall English-language usage. Significant number of the participant students reported devoting more effort to studying the language elements that frequently appear on the ESSLCEE. The FGD report has a positive relationship with the survey data that suggests how the high-stakes ESSLCEE has affected students’ attempts to strengthen their command of English during their independent study periods.

The FGD made regarding reading comprehension questions provided additional proof those students are not utilizing the language development skills as much as is necessary. One participant complained that he did not have access to the reading texts for the ESSLCEE and hence could not practice them as he does for other language components. The data back up the earlier findings that Allen [1], Sato [81], and Shih [6] reported.

In short, the data from the questionnaire, FGD, and content analysis suggested that students practice language components connected to the ESSLCEE more frequently than those unrelated to the ESSLCEE. The outcome suggests that the exam has a negative washback on each student’s personal effort to develop their communication abilities. The results are in line with those of earlier research projects carried out by Mickan and Motteram [5] and Zhan and Andrews [12].

4.2. Significance Differences among Students across Their Groups in Practicing Non-ESSLCEE-Related Language Components Out-of-Classroom. The aim of this section is to answer the second research question which seeks to check if the differences among the students with their academic achievement levels in practicing (studying) the non-ESSLCEE-related English activities out of the classroom are significant (Table 6).

As can be shown in Table 6, among the types of non-ESSLCEE-related language components practiced out-of-classroom, the total mean value scored within grammar activities is $M = 3.34$ which is the highest mean score when compared to the responses given to the remaining non-ESSLCEE-related items. In the contrary, the least mean value ($M = 2.1$) is registered within practicing of reading comprehensions activities. As the statistical reports registered by different achieving groups in responses to the all non-ESSLCEE-related items show, students in “Excellent” achieving

TABLE 7: Mean differences of students' responses to non-ESSLCEE-related items.

	N	Mean	SD	Std. error	95% Confidence interval for mean		Mini	Max
					Lower bound	Upper bound		
Fair (50–59)	25	2.31	0.182	0.036	2.24	2.39	2	3
Satisfactory (60–79)	43	2.31	0.222	0.034	2.25	2.38	2	3
Very good (80–89)	19	3.54	0.233	0.053	3.42	3.65	3	4
Excellent (90–100)	7	3.67	0.289	0.109	3.40	3.93	3	4
Total	94	2.66	0.606	0.062	2.54	2.79	2	4

groups registered the highest total mean value ($M=3.67$) when compared to others. The “Fair” and “Satisfactory” achieving groups, however, scored the least mean value ($M=2.31$). The responses given to each of the non-ESSLCEE-related items by each of the individual achieving groups conform to the observed reports of the cumulative mean value. For instance, in practicing English oral conversations with friends, “Fair” and “Satisfactory” achieving group registered the mean value of 2.48 and 2.53 whereas “Very Good” and “Excellent” achieving group registered the mean value of 3.32 and 3.43, respectively. Similarly, “Fair” ($M=1.72$) and “Satisfactory” ($M=1.65$) achieving groups registered least mean value in response to practicing reading texts from various authentic materials. This can be understood by comparing the mean scores against of the “Very Good” ($M=3.05$) and the “Excellent” ($M=3.57$) achieving groups. The statistical results show that the high-achieving groups (“Very Good” and “Excellent”) spend more time in practicing non-ESSLCEE-related language components out-of-classroom than the low-achieving groups (“Fair” and “Satisfactory”).

As can be understood from the table (Table 6), the “Excellent” and the “Very Good” achievers took more time than did the “Satisfactory” and the “Fair” achieving groups to practicing non-ESSLCEE-related language components for improvement their English. The finding may be related with the learning hypotheses formulated by Alderson and Wall [23] and study conducted by Pan (40). On the other hand, almost all students in different achievement groups spend more time in practicing non-ESSLCEE-related grammar and vocabulary activities than the rest of the language skills. This finding can also be connected with the past findings of Pan [82].

It is noted that the registered mean value vary across all groups. For further clarification, see Table 7 below.

As shown in Table 7, among all the four categories of academic achievements “Fair” (50–59), “Satisfactory” (60–79), “Very Good” (80–89), and “Excellent” (90–100), there were mean differences (2.31, 2.34, 3.48, and 3.67, respectively) on the time they invest practicing non-ESSLCEE-related language components. The data display that the non-ESSLCEE-related activities were most frequently practiced by the “Excellent” scorer groups. However, the “Fair” and “Satisfactory” scorer groups registered within the least frequency. Thus, to check whether their differences across the groups are significant, the mean score was computed using one-way ANOVA.

TABLE 8: ANOVA test of significance difference of non-ESSLCEE-related items.

	Sum of squares	df	Mean square	F	Sig.
Between groups	29.797	3	9.932	206.000	0.000
Within groups	4.339	90	0.048		
Total	34.136	93			

As can be understood from the table (Table 8), the observed differences among the four groups are statistically significant. That is the significance value of differences among students of different academic achievements levels regarding the practices they make on non-ESSLCEE-related items out of the classrooms is not greater than 0.05: $F(3,93) = 0.000, p < 0.05$.

To verify which pair of means significantly differed, post hoc comparisons conducted using Tukey HSD test. The test revealed that the mean scores for the “Fair” achieving group ($M=2.31, SD=0.182$) and “Satisfactory” achieving group ($M=2.31, SD=0.222$) were significantly different from (a) “Very Good” achieving group ($M=3.54, SD=0.233$) and (b) “Excellent” achieving group ($M=3.67, SD=0.289$). However, there was no significant difference between “Fair” ($M=2.31, SD=0.182$) and “Satisfactory” ($M=2.31, SD=0.222$) achieving groups, and similar absence of significant difference is also observed between the “Very Good” ($M=3.54, SD=0.233$) and “Excellent” ($M=3.67, SD=0.289$) achieving groups. The comparisons show that the low-achieving groups (“Fair” and “Satisfactory”) spend less time of studying non-ESSLCEE-related language components than the high (“Very Good” and “Excellent”) achieving groups. This finding proves the learning hypotheses formulated by Alderson and Wall [23] and partially confirms the studies conducted by Allen [1]. However, the finding is in contrast with Pan’s [11], Buyukkeles’ [10], Cheng et al.’s [35], and Shohamy et al.’s [24] with the possible reasons (see the discussion). Another important reason for the variations of washback of test on students’ leaning is the matter of context (in class or out of class) in which students practice [5, 12, 50].

4.3. Differences among Students across Their Groups in Practicing ESSLCEE-Related Language Components Out of the Classroom. Here, the section is aimed to address the last research question to identify the degree of differences among students of different academic achievement levels in

TABLE 9: Comparative analyses of ESSLCEE-related items across the groups.

Items	Fair (50–59)	Satisfactory (60–79)	Very good (80–89)	Excellent (90–100)	Total
Practicing ESSLCEE-related reading comprehensions	3.92	3.93	3.68	4	3.88
Practicing ESSLCEE-related vocabulary questions	4.16	4.12	4.26	4.43	4.18
Practicing ESSLCEE-related grammar questions	4.36	4.49	4.53	4.86	4.49
Practicing ESSLCEE-related dialog questions	4.24	4.28	4.37	4	4.27
Practicing ESSLCEE-related writing questions	4.08	4.35	4.42	4.29	4.29
Total	4.15	4.23	4.25	4.26	4.22

TABLE 10: Mean differences of students responses to ESSLCEE-related items.

	N	Mean	SD	Std. error	95% Confidence interval for mean		Mini	Max
					Lower bound	Upper bound		
Fair (50–59)	25	4.15	0.176	0.035	4.08	4.22	4	4
Satisfactory (60–79)	43	4.23	0.218	0.033	4.17	4.30	4	5
Very good (80–89)	19	4.25	0.239	0.055	4.14	4.37	4	5
Excellent (90–100)	7	4.26	0.190	0.072	4.08	4.43	4	5
Total	94	4.22	0.211	0.022	4.17	4.26	4	5

studying (practicing) ESSLCEE-related learning activities out of the classroom. The responses of the students to the items 7–11 are analyzed comparatively (Table 9).

As can be seen from Table 9, the highest cumulative mean score ($M = 4.49$) was registered in practicing ESSLCEE-related grammar questions regardless of the mean differences exist among each of the achieving groups. Conversely, least cumulative mean score ($M = 3.88$) was documented in practicing ESSLCEE-related reading comprehension questions. When the total mean scores of the responses given to ESSLCEE-related items by each of the achieving groups are compared, it is observed that “Fair” achieving groups scored the least mean value ($M = 4.15$) whereas the “Excellent” achieving groups scored the highest mean value ($M = 4.26$). On the other hand, among the responses given to the ESSLCEE-related items, the highest mean value ($M = 4.86$) was registered by the “Excellent” achieving group to the response of practicing ESSLCEE-related grammar questions. Nonetheless, the least mean value ($M = 3.68$) was registered by the “Very Good” achieving group in response to practicing ESSLCEE-related reading comprehension questions. The finding confirmed the study conducted by Allen [1], Sato [81], and Shih [6].

However, mean differences are observed across all the four groups (Table 10).

As it can be observed from the table (Table 10), among all the four categories of academic achievements “Fair” (50–59), “Satisfactory” (60–79), “Very Good” (80–89), and “Excellent” (90–100), there were mean differences (4.15, 4.23, 4.25, and 4.26, respectively) on the time they spend practicing ESSLCEE-related language skills. The cumulative mean score of each groups show that “Very Good” and “Excellent” scorers more frequently practice the ESSLCEE-related activities than the rests. To check whether their differences across the groups are significant or not, the mean score was computed using

TABLE 11: ANOVA test of significance difference of ESSLCEE-related items.

	Sum of squares	df	Mean square	F	Sig.
Between groups	0.151	3	0.050	1.141	0.337
Within groups	3.981	90	0.044		
Total	4.133	93			

one-way ANOVA [83]. To conduct the test, assumptions of parametric test statistics were computed and met [84].

From the table below (Table 11), it can be understood that there is no statistically significant difference between the different achievement groups of students’ self-reported practices of ESSLCEE-related items out-of-classrooms. This happened albeit the students’ academic achievement differences (because the significance value is greater than 0.05), $F(3, 93) = 0.337$, $p > 0.05$. Though there were mean differences among students of different academic achievement levels in their practicing of ESSLCEE-related learning activities out of classrooms, the differences are insignificant. The variation of the current finding with the previous studies [10] might be due to various reasons [23, 33, 34, 38, 50].

One thing to be noted here is that the data from Tables 6 and 9 show as the “Fair” and “Satisfactory” achieving groups registered high mean score in responses to ESSLCEE-related items, but low means scores responses to non-ESSLCEE-related items. However, “Very Good” and “Excellent” achieving groups registered high mean score in both non-ESSLCEE-related and ESSLCEE-related items than the “Fair” and “Satisfactory” achieving groups. This finding is supported by [11, 28, 68].

Overall, the data from the table demonstrate that, when compared to the non-ESSLCEE-related items (Table 6), students in all achievement groups registered high mean scores

for all ESSLCEE-related items (Table 9). Regardless of their categories of academic achievement, the results reveal that students spend more time practicing the language components linked to the ESSLCEE than the language components unrelated to the ESSLCEE. The findings may support earlier research by Buyukkeles [10] and Pan [11].

5. Discussion

The primary goal of this study is to examine the influence of high-stakes ESSLCEE on students' learning determining to what extent they practice ESSLCEE- and non-ESSLCEE-related learning activities out-of-classroom. According to the data from Table 3, students reported practicing ESSLCEE-related language components more frequently than non-ESSLCEE-related ones during their study time. The results showed that students were more interested in studying the language parts of the exams during their after-class English study time. The results are in line with earlier research done by Mickan and Motteram [5] and Zhan and Andrews [12]. Zhan and Andrews [12] examined the extent to which the revised College English Test Band 4 (CET-4) actually influenced Chinese non-English-major undergraduates' out-of-class learning. According to the findings, students were more likely to alter what they learned than how they learned when the goal test was present. According to Mickan and Motteram [5], the candidates concentrated on the test-preparation activities, which had a beneficial washback on the students' English learning. In contrast to Mickan and Motteram's findings, the high-stakes ESSLCEE had a detrimental impact on the students' out-of-classroom English-learning habits in the current study.

It was found that students tended to focus on tasks and materials that related to the test when preparing for it. On the other hand, the students also appeared to place less emphasis on using language skills that are essential to their lives but are excluded from or given less weight in the ESSLCEE. They frequently concentrate more on the writing, grammar, and dialog sections of the exam. The effects of the high-stakes test on students' learning are apparent when the data from FGD and the mean scores recorded with both non-ESSLCEE-related and ESSLCEE-related items (Table 3) are compared with the percentage composition of language components found in previous ESSLCEE questions (Table 5). For instance, listening is one of the receptive skills that is not tested in the exam and is also the skill that students with the lowest mean score ($M = 2.15$) in non-ESSLCEE-related items exercise the least. Second, reading comprehension and vocabulary exercises are the two skills that have received the least weight in previous ESSLCEE questions. Reading comprehension is also mentioned last in the list of talents. On the other hand, vocabulary is listed in the second rank order with its mean score under both non-ESSLCEE-related and ESSLCEE-related categories, and its outcome is related to the findings of Pan and Newfields [28] and Pan [82]. According to the experts, the amount of vocabulary included in exam content has a negligible impact on how students learn language. Students are aware of studying vocabulary and grammar even when there is no relation to a high-stakes

exam because they believe that doing so would help them become more fluent in English [82]. Third, one of the FGD participants explained that he uses his spare time after class to study exam-related information and neglects non-ESSLCEE-related language contents due to time constraints. The lack of access to reading comprehension questions for the ESSLCEE was mentioned by another FGD participant. The results can be connected to earlier results from Allen [1], Sato [81], and Shih [6]. According to Sato [81], resource limitations were one of the elements influencing participant learning behavior. When they do not have a companion with whom to practice speaking, pupils avoid studying for speaking tests, according to a comparable case study by Allen and Sato. Because they do not have access to the resources, the students in the current study appear to avoid practicing reading comprehension questions. As a result, the evidence derived from various data points to the likelihood that students will modify their outside-of-classroom English study time in accordance with the volume of language components included in previous ESSLCEE questions. According to the data, regardless of students' differences in academic achievement, high-stakes English tests have a greater negative impact on their autonomous language learning. This implies that how many language components are included in the past ESSLCEE questions affects how much time students spend on each language component. Less time may be spent studying if the exam's language components are given less weight. The discovery supports earlier research [12, 19, 63, 64, 81]. For instance, tests will affect the rate and sequence, degree, and depth of learning, according to Alderson and Wall's theories. The current study, however, differs from previous research in that it examined the impact of high-stakes testing on students' outside-of-classroom learning context.

Testing the differences in non-ESSLCEE-related learning activities across students with varying academic success levels is the second and main goal of this study. Students of varying academic achievement levels exercised the language components outside of the classroom in various ways, as seen in Table 6. The "Fair" and "Satisfactory" achieving groups practiced non-ESSLCEE-related language components less frequently than the "Excellent" and "Very Good" achieving groups. Table 6 shows that the Post Hoc Tests results demonstrate that there is no significant difference in the amount of time students in low-achieving groups ("Fair" and "Satisfactory") spend outside of class studying the test-independent English-learning activities. Similarly, when it comes to applying the learning activities, there is no discernible difference between the groups that get "Excellent" and "Very Good." The distinction between "Fair" achievers and "Very Good" and "Excellent" achievers, as well as between "Satisfactory" achievers and "Very Good" and "Excellent" achievers, is, however, rather evident. Although not enough, the FGD discussion results among students do confirm the findings of the statistical tests. One of the interviewees said he had never budgeted his study time for language learning activities unrelated to exams.

The results support Alderson and Wall's hypotheses [19] about how certain students will experience washback on tests but not others. The research supported the conclusions made

by Pan [11] and Buyukkeles [10]. When preparing for high-stakes tests, low-proficiency students did not spend as much time on specific types of language-skill developing activities as intermediate and high-proficiency students, according to Pan [11]. Buyukkeles' [10] research also revealed that some students may not feel the need to work on their language skills, which have insignificant contribution in high-stakes tests. Pan [82] provided some explanations for why all the high-achieving groups worked vocabulary and grammar drills unrelated to the ESSLCEE more than they did other language skills. She asserted that students can focus more on grammar and vocabulary exercises than on other language skills because they believe these subskills will help them become more fluent in English. The information gathered from the FGD participants supports the assertion. They claimed that they spent their time outside of the classroom studying grammar and vocabulary in the hopes that their command of the English language would improve. However, the finding that supports the research done by Allen [1] is the significant difference in the frequency of non-ESSLCEE-related activities between the high-achieving groups ("*Excellent*" and "*Very Good*") and low-achieving groups ("*Fair*" and "*Satisfactory*"). The amount of fluency practice students put in for IELTS exams varies significantly amongst students in different achieving groups, according to the author.

The result, however, disagrees with those of Pan [11], Buyukkeles [10], Cheng [35], Pan [11], Shohamy et al. [24], and Pan. The disparity between these findings and those of previous studies may be due to differences in the way that students were divided into academic levels, differences in the nature or status of the test—national (ESSLCEE) versus international tests (IELTS and TOFEL), or differences in the homogeneity or heterogeneity of students—students from different context or from similar context. The variance in results may not come as a surprise given the complexity of the washback research [19, 33, 34, 38, 50]. In addition, it is asserted that learners' washback on learning is not consistent. Furthermore, the environment in which students live and struggle to prepare for exams is a significant factor in the learning differences that may exist among students. Because of this, the impact of the exam on students' learning may differ significantly from that of the classroom context. According to research, students who study independently outside of class and tend to concentrate on tasks and materials related to the test can adopt more diverse preparation strategies than is likely to happen in a classroom setting [5, 12, 50].

Examining the differences in how students with various academic achievement levels approach learning ESSLCEE-related language components is another important goal of the study. Table 9 demonstrates that, with the exception of reading comprehension questions, all academic achievement groups spent a significant amount of their out-of-classroom learning time on all ESSLCEE-related language components. Surprisingly, ESSLCEE-related reading texts received less attention from all the high-achieving groups than the other language skills. Table 8 shows that there is no statistically significant difference in students' out-of-classroom learning methods for language components relevant to the ESSLCEE

amongst the various achieving groups. The current finding is in opposition to earlier studies by Pan [11] and Buyukkeles [10]. Several distinct reasons can be to blame for the discrepancies in the results. The participants were told via various social media platforms that the exam administration system will transition from paper to online, which is the only compelling argument for the difference in results. Consequently, all the candidates—including teachers, parents, and school administrators—were extremely enthusiastic and sought to pave the way for success in the exam by setting up frequent tutorial classes, preparing model questions, forming teams, and assigning better achiever students to each team so they could work together to practice test-related questions both inside and outside of school grounds. As a result, the candidates' willingness to work together to share information and experiences regarding the upcoming test caused them to engage in ESSLCEE-related activities on a regular basis, which may have led to minimal differences among students.

The comparison of Table 9 result with Table 6's reveals that both "*Fair*" and "*Satisfactory*" achieving groups are more affected by the washback of ESSLCEE than the "*Very Good*" and "*Excellent*" achieving groups. However, the "*Very Good*" and "*Excellent*" achieving groups showed higher mean scores in both non-ESSLCEE-related and ESSLCEE-related items when compared to the "*Fair*" and "*Satisfactory*" achieving groups. The investigations undertaken by Ferman [68], Pan [11], and Shohamy et al. [24] provide evidence for the relative consistency of the high-achieving groups ("*Very Good*" and "*Excellent*"). The authors reported that high scorers are consistently willing to learn independent of the push of the high-stakes tests. According to Pan [11], higher academically performing groups embraced a variety of test-related and linguistic skill-building activities. According to Ferman [67] and Shohamy and his colleagues [24], students who performed better academically were more interested in studying both language skills needed for exams and language skills not required for exams.

The study's conclusions, which were drawn from all the data analyzed, appear to corroborate two main assertions. One is the fact that during their individualized out-of-classroom English study sessions, students in diverse academic achievement groups display washback variability. The second finding is that students spend more time working on ESSLCEE-related learning activities out-of-classroom than they do on language components that are not related to ESSLCEE, despite differences in academic achievement. This clearly shows that the washback of high-stake ESSLCEE has a detrimental effect on students' attempts to learn English out-of-classroom, regardless of the specific academic achievement group to which they belong. The results might lend support to past studies by Pan [11] and Buyukkeles [10]. In addition, the findings can provide light on the Alderson and Wall washback theory, which predicts that "tests will have washback on all or some learners."

6. Conclusion

The current study represents the first attempt to examine how, in an Ethiopian context, ESSLCEE influences students'

autonomous learning habits out-of-classroom. Because it is one of the few studies that concentrated on the washback of testing-related learners' viewpoints, the study believed that it would better understanding of learning washback on how tests affect students' independent out-of-classroom learning practices. In addition, it highlights the extent to which the high-stakes tests have an impact on students across a range of academic achievement levels.

The study found that students from low-achieving groups ("Fair" and "Satisfactory" scorers) and high-achieving groups ("Very Good" and "Excellent" scorers) significantly differed in their ability to independently study non-ESSLCEE-related language components out-of-classroom. Because of their academic backgrounds, high-stakes test takers appear to have experienced a negative washback of variability. It has been found that students in low-achieving groups do not devote as much time to specific language-communicative skill-building activities as students in high-achieving groups do. The result validates Alderson and Wall's hypothesis about how assessments affect the rate, sequence, degree, and depth of learning and how some learners will experience washback but not others.

The study also reveals that there is no statistically significant difference between low- and high-achieving groups in terms of independent out-of-classroom activities for language components relevant to ESSLCEE. The high-achieving groups are less impacted by the washback of high-stakes ESSLCEE because they spend more time practicing ESSLCEE-related and non-ESSLCEE-related language components than the low-achieving groups. In overall, the study undoubtedly proves the ESSLCEE's washback variability on the efforts students make to learn English outside of class, depending on the specific academic achievement group to which they belong.

7. Implication of the Findings

Strong evidence was discovered, as was already mentioned, that high-stakes ESSLCEE had a detrimental effect on students' out-of-classroom English learning habits. Additionally, students with varying degrees of academic achievement levels have a sizable amount of washback variability. There is obviously washback from the high-stakes ESSLCEE for all four achieving groups, but it is more pronounced for the low achieving groups ("Fair" and "Satisfactory" scorers). The study has implications that EFL teachers should put in place various mechanisms to encourage positive washback, commencing before the start of the classroom lesson. In order for students to be aware of the use of the target language, EFL teachers should inform the students about how English will be used in the future in their programs. The development of language skills depends heavily on placing students in circumstances where they become aware of out-of-classroom language learning practices. The idea of "learner autonomy" should be communicated to students, and strategies for fostering it should be recommended. EFL teachers should assist students in identifying their preferred strategies of learning English and then instruct them on how to apply those

strategies both in and out of the classroom. In addition, they are urged to support their students in using strategies that help them practice each language element equally. The discovery also implies that the test design for the EFL should be reviewed. The test's weighting should be in line with the curriculum. In other words, the test elements should unquestionably be equally weighted. Students will be expected to give each component equal weight in this manner. In this way, students would be expected to attach equal importance to each component. The test should be more communicative in its nature. The test content could be redesigned to achieve this. It would be wise to test students' command of the English language directly or by seeing how well they produce language rather than using discrete objective items. The focus would shift from linguistic accuracy to communicative competence, potentially altering how students learn. The findings of this study can also provide policy makers, syllabus designers, educators, and other stakeholders with insights on how to promote positive washback.

8. Limitations and Future Trajectories

When conducting research, limitations are a given. As a result, there are several limitations for the current study that should be taken into account for subsequent research in the field. First, the study concentrated on the context of secondary school students in Ethiopia, which might not accurately reflect the situation of students in other countries. Future scholars can therefore expand their investigation into areas that the current study did not cover. Second, all of the survey questions were of the closed-ended variety. More closed and open-ended items would need to be added to a future study of a similar nature to get more detailed data. Third, rather than delivering the self-reported items to the respondents all at once, the participants should be required to keep a written record every day for a specific number of weeks to get more accurate data on what students independently practice outside of the classroom. Finally, future researchers should either develop distinct FGD based on their achievement groups or conduct interviews separately to avoid the over dominance of some participants on the others during FGD.

Data Availability

The data that support the findings of this study are available from the corresponding author, (Getachew Desalegn Debisa), upon reasonable request.

Conflicts of Interest

The authors declare that there are no conflicts of interest in conducting this research.

Acknowledgments

We would like to extend our heartfelt gratitude to participants of the study and administrative bodies of Leka-Nekemte and Arjo Secondary Schools for their cooperation during the study. This study was financially

supported by Wollega University, Nekemte, Ethiopia with Grant Number (WUSGS/207/2012).

References

- [1] D. Allen, "Investigating washback to the learner from the IELTS test in the Japanese tertiary context," *Language Testing in Asia*, vol. 6, no. 1, Article ID 20, 2016.
- [2] M. Dong and X. Liu, "Impact of learners' perceptions of a high-stakes test on their learning motivation and learning time allotment: a study on the washback mechanism," *Heliyon*, vol. 8, no. 12, Article ID e11910, 2022.
- [3] C. M. Gosa, "Investigating washback: a case study using student diaries," PhD dissertation, Lancaster University, 2004.
- [4] N. Mahmud, N. H. M. Lateh, N. Mahmud, A. A. Hassan, A. F. Mohamed, and S. A. A. Tarmizi, "Washback impact of the MUET: the before and after effect of a high-stake University English Test in Malaysia," *International Journal of Learning, Teaching and Educational Research*, vol. 20, no. 8, pp. 1–17, 2021.
- [5] P. Mickan and J. Motteram, "The preparation practices of IELTS candidates: case studies," *IELTS Research Reports*, vol. 10, pp. 223–262, 2009.
- [6] C.-M. Shih, "A new washback model of students' learning," *Canadian Modern Language Review*, vol. 64, no. 1, pp. 135–161, 2007.
- [7] C. L. Tsang, "Examining washback on learning from a sociocultural perspective: the case of a graded approach to english language testing in Hong Kong," MA dissertation, 2017.
- [8] Q. Xie and S. Andrews, "Do test design and uses influence test preparation? Testing a model of washback with structural equation modeling," *Language Testing*, vol. 30, no. 1, pp. 49–70, 2013.
- [9] D. Tsagari, "Revisiting the concept of test washback: investigating FCE in Greek language schools," *Cambridge ESOL, Research Notes*, vol. 35, pp. 5–9, 2009.
- [10] G. Buyukkeles, "The washback effect of a high-stakes exit test on students' motivation in a Turkish pre-university EFL preparatory school," MA dissertation, https://www.teachingenglish.org.uk/sites/teacheng/files/mda2017_university_of_reading_guliz_buyukkeles.pdf, 2014.
- [11] Y. Pan, "Learner washback variability in standardized exit tests," *The Electronic Journal for English as a Second Language: ESL-EJ*, vol. 18, no. 2, pp. 1–30, 2014.
- [12] Y. Zhan and S. Andrews, "Washback effects from a high-stakes examination on out-of-class English learning: insights from possible self-theories," *Assessment in Education: Principles, Policy and Practice*, vol. 21, no. 1, pp. 71–89, 2014.
- [13] J. Heaton, "Classroom Testing," Longman, London, 1990.
- [14] M. S. Islam, M. K. Hasan, S. Sultana, A. Karim, and M. M. Rahman, "English language assessment in Bangladesh today: principles, practices, and problem," *Language Testing in Asia*, vol. 11, no. 1, pp. 2–21, 2021.
- [15] H. Salehi and MdYunus, "The washback effect of the Iranian universities entrance exam: teachers' insights," *Journal of Language Studies*, vol. 12, no. 2, pp. 609–628, 2012.
- [16] N. Beikmahdavi, "Washback in language testing: Review of related literature first," *International Journal of Modern Language Teaching and Learning*, vol. 1, no. 4, pp. 130–136, 2016.
- [17] L. Cheng, *Changing Language Teaching through Language Testing: A Washback Study*, Cambridge University Press, Cambridge, 2005.
- [18] N. B. Manjarres, "Washback of the foreign language test of the state examinations in Colombia: a case study," *Arizona Working Papers in SLAT*, vol. 12, pp. 1–19, 2005.
- [19] J. C. Alderson and D. Wall, "Does washback exist?" *Applied Linguistics*, vol. 14, no. 2, pp. 115–129, 1993.
- [20] K. M. Bailey, "Working for washback: a review of the washback concept in language testing," *Language Testing*, vol. 13, no. 3, pp. 257–279, 1996.
- [21] S. Gates, "Exploiting washback from standardized tests," in *Language Testing in Japan*, J. D. Brown and S. O. Yamashita, Eds., pp. 101–106, Japanese Association, Tokyo, 1995.
- [22] A. Hughes, *Testing for Language Teachers*, Cambridge University Press, 3rd edition, 2003.
- [23] S. Messick, "Validity and washback in language testing," *Language Testing*, vol. 13, no. 3, pp. 241–256, 1996.
- [24] E. Shohamy, S. Donitsa-Schmidt, and I. Ferman, "Test impact revisited: washback effect over time," *Language Testing*, vol. 13, no. 3, pp. 298–317, 1996.
- [25] J. D. Brown and K. M. Bailey, "Language testing courses: what are they in 2007?" *Language Testing*, vol. 25, no. 3, pp. 349–383, 2008.
- [26] S. Messick, "Meaning and values in test validation: the science and ethics of assessment," *Educational Researcher*, vol. 18, no. 2, pp. 5–11, 1989.
- [27] N. Mahmud, "Investigating the washback effect of the MUET as a University entry test on students in Malaysia," PhD dissertation, <https://core.ac.uk/download/pdf/196513711.pdf>, 2018.
- [28] Y.-C. Pan and T. Newfields, "Tertiary EFL proficiency graduation requirements in Taiwan: a study of washback on learning," *Electronic Journal of Foreign Language Teaching*, vol. 9, no. 1, pp. 108–122, 2012.
- [29] P.-C. Shih, "The english benchmark policy for graduation: an investigation of perception, motivation, and approaches to learning at a University of Technology in Central Taiwan," PhD dissertation, http://etheses.dur.ac.uk/8512/1/Pei-Chi_Shih.pdf?DDD29%2B.pdf, 2013.
- [30] B. W. Stoneman, "The impact of an exit English test on Hong Kong undergraduates: a study investigating the effects of test status on students' test preparation behaviors," PhD dissertation, 2006.
- [31] D. Tsagari and L. Cheng, "Washback, impact, and consequences revisited," in *Language Testing and Assessment*, E. Shohamy, I. G. Or, and S. May, Eds., pp. 359–372, Springer, Cham, Switzerland, 3rd edition, 2017.
- [32] D. Brown, *Language Assessment—Principles and Classroom Practice*, Pearson Education, White Plains, NY, USA, 2004.
- [33] J. C. Alderson, "Foreword," in *Washback in Language Testing: Research Contexts and Methods*, L. Cheng, Y. Watanabe, and A. Curtis, Eds., pp. ix–xii, Routledge, 2004.
- [34] J. C. Alderson and L. Hamp-Lyons, "TOEFL preparation courses: a study of washback," *Language Testing*, vol. 13, no. 3, pp. 280–297, 1996.
- [35] L. Cheng, A. Andrews, and Y. Yu, "Impact and consequences of school-based assessment (SBA): students' and parents' views of SBA in Hong Kong," *Language Testing*, vol. 28, no. 2, pp. 221–249, 2011.
- [36] A. Gebril and G. T. Brown, "The effect of high-stakes examination systems on teacher beliefs: Egyptian teachers' conceptions of assessment," *Assessment in Education: Principles, Policy and Practice*, vol. 21, no. 1, pp. 16–33, 2014.
- [37] N.-H. Luong-Phan and G. Effeney, "TOEFL iBT and language learning motivation: an investigation into teaching styles and influential factors for Vietnamese adolescents," *International*

- Journal of Research Studies in Language Learning*, vol. 4, no. 3, pp. 3–18, 2015.
- [38] L. Cheng, “The key to success: english language testing in China,” *Language Testing*, vol. 25, no. 1, pp. 15–37, 2008.
- [39] M. Spratt, “Washback and the classroom: the implications for teaching and learning of studies of washback from exams,” *Language Teaching Research*, vol. 9, no. 1, pp. 5–29, 2005.
- [40] M. Estaji and Z. Tajeddin, “The learner factor in washback context: an empirical study investigating the washback of the IELTS academic writing test,” *Language Testing in Asia*, vol. 2, no. 1, pp. 5–25, 2012.
- [41] L. Taylor, “Washback and impact,” *ELT Journal*, vol. 59, no. 2, pp. 154–155, 2005.
- [42] Z. W. Ayele, “Wash-back effects of national higher education entrance examinations on students’ english language learning: in selected Schools of North Western Ethiopia,” *International Journal of African and Asian Studies*, vol. 46, no. 4, pp. 28–33, 2018.
- [43] S. Gashaye, “The washback of the Ethiopian National English Examination (ENEE) for grade 10 on students’ practice: Debre Markos secondary schools on focus,” *Theory and Practice in Language Studies*, vol. 11, no. 7, pp. 780–787, 2021.
- [44] E. K. Reta, “The impact of high stake EFL tests on the role of teachers, learners and parents,” *The International Journal of Research in Teacher Education*, vol. 9, no. 3, pp. 1–8, 2018.
- [45] T. McNamara, *Language Testing*, Oxford University Press, Oxford, 2000.
- [46] L. Cheng and A. Curtis, *Washback or backwash: A review of the impact of testing on teaching and learning: Washback in language testing: Research contexts and methods*, Lawrence Erlbaum Associates, Mahwah, NJ, 2004.
- [47] K. Morrow, “The evaluation of tests of communicative performance,” in *Innovations in Language Testing*, M. Portal, Ed., pp. 1–3, Routledge, London, 1986.
- [48] W. J. Popham, “The merits of measurement-driven instruction,” *Phi Delta Kappa*, vol. 68, pp. 679–682, 1987.
- [49] G. F. Madaus, “The influence of testing on the curriculum,” in *Critical Issues in Curriculum: Eighty-seventh Yearbook of the National Society for the Study of Education*, L. N. Tanner, Ed., pp. 83–121, University of Chicago Press, Chicago, 1988.
- [50] A. Green, *IELTS Washback in Context*, Cambridge University Press, Cambridge, 2007.
- [51] A. Hughes, Backwash, and TOEFL, “Backwash and TOEFL 2000,” University of Reading, Unpublished manuscript, 2000.
- [52] S. Saif, “Theoretical and empirical considerations in investigating washback: a study of ESL/EFL learners,” PhD dissertation, 2006.
- [53] Y. Watanabe, “Methodology in washback studies,” in *Washback in Language Testing: Research Contexts and Methods*, L. Cheng, Y. Watanabe, and A. Curtis, Eds., pp. 19–36, Routledge, 1st edition, 2004.
- [54] D. Wall, *The Impact of High-stakes Examinations on Classroom Teaching: A Case Study using Insights from Testing and Innovation Theory*, Cambridge University Press, 2005.
- [55] D. Wall, “The impact of high stakes testing on teaching and learning: can this be predicted or controlled?” *System*, vol. 28, no. 4, pp. 499–509, 2000.
- [56] H. Y. Chu, “Stakes, needs and washback: an investigation of the English benchmark policy for graduation and EFL education at two technological universities in Taiwan,” PhD dissertation, National Taiwan Normal University, 2009.
- [57] H. Li, Q. Zhong, and H. K. Suen, “Students’ perceptions of the impact of the college english test,” *Language Testing in Asia*, vol. 2, no. 3, pp. 77–94, 2012.
- [58] E. Smyth and J. Banks, “High stakes testing and student perspectives on teaching and learning in the Republic of Ireland,” *Educational Assessment, Evaluation and Accountability*, vol. 4, no. 4, pp. 283–306, 2012.
- [59] Y. Tsai and C. H. Tsou, “A standardised English language proficiency test as the graduation benchmark: student perspectives on its application in higher education,” *Assessment in Education: Principles, Policy and Practice*, vol. 16, no. 3, pp. 319–330, 2009.
- [60] Y. Watanabe, “Does the university entrance examination motivate learners? A case study of learner interviews,” in *Akita Association of English Studies, Trans-equator Exchanges: A Collection of Academic Papers in Honour of Professor David Ingram*, pp. 100–110, 2001.
- [61] M. Dong, “Do school status and classroom English learning practices influence Chinese learners’ perceptions of a high-stakes English test?” *Language Testing in Asia*, vol. 12, no. 1, pp. 1–14, 2022.
- [62] A. Stopar and G. Ilc, “Reading for a test: the effect of high-stakes exams on reading strategies,” *Porta Linguarum Revista Interuniversitaria de Didáctica de las Lenguas Extranjeras*, pp. 103–115, 2017.
- [63] Q. Xie, “Do component weighting and testing method affect time management and approaches to test preparation? A study on the washback mechanism,” *System*, vol. 50, pp. 56–68, 2015.
- [64] Y. Zhan and Z. H. Wan, “Test takers’ beliefs and experiences of a high-stakes computer-based English listening and speaking test,” *RELC Journal*, vol. 47, no. 3, pp. 363–376, 2016.
- [65] J. Fan, P. Ji, and X. Song, “Washback of university-based English language tests on students’ learning: a case study,” *The Asian Journal of Applied Linguistics*, vol. 1, no. 2, pp. 178–191, 2014.
- [66] J. Fan and P. Ji, “Test candidates’ attitudes and their test performance: the case of the Fudan English Test,” in *University of Sydney, Papers in TESOL*, vol. 9, pp. 1–35, 2014.
- [67] D. Tsagari, “Review of washback in language testing: what has been done? what more needs doing?” 2007, <https://www.semanticscholar.org/paper/Review-of-Washback-in-Language-Testing%3A-What-Has-Tsagari/21c0df7f2e5d62dab2c6910a45a1a145b9fb28ee>.
- [68] I. Ferman, “The washback of an EFL national oral matriculation test to teaching and learning,” in *Washback in Language Testing, Research Contexts and Methods*, L. Cheng, Y. Watanabe, and A. Curtis, Eds., pp. 191–210, Lawrence Erlbaum Associates, Mahwah, NJ, USA, 2004.
- [69] L. Chala and M. Agago, “Exploring national examination malpractice mechanisms and countermeasures: an Ethiopian perspective,” *International Journal of Instruction*, vol. 15, no. 3, pp. 413–428, 2022.
- [70] T. Gashaw, “The measurement of english language proficiency of high school graduates,” MA thesis, 1982.
- [71] D. Leta and Achievement, “Washback and proficiency in school leaving examination: a case of innovation in an Ethiopian setting,” Unpublished PhD dissertation, Lancaster University, 1990.
- [72] J. W. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods and Approaches*, SAGE Publications, Thousand Oaks, CA, USA, 4th edition, 2014.
- [73] J. A. Maxwell, “Expanding the history and range of mixed methods research,” *Journal of Mixed Methods Research*, vol. 10, no. 1, pp. 12–27, 2016.

- [74] C. Poth and S. E. P. Munce, "Commentary—preparing today's researchers for a yet unknown tomorrow: promising practices for a synergistic and sustainable mentoring approach to mixed methods research learning," *International Journal of Multiple Research Approaches*, vol. 12, no. 1, pp. 56–64, 2020.
- [75] J. W. Creswell and J. D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, SAGE Publications Inc, 5th edition, 2018.
- [76] M. M. Bergman, *Advances in Mixed Methods Research*, SAGE Publications Ltd, 2008.
- [77] MOE, *Final Report Study on Medium of Instructions in Primary Schools in Ethiopia*, Addis Ababa, 2007.
- [78] Y. S. Kumar, *Fundamental of Research Methodology and Statistics*, New Age International (P) Ltd, New Delhi, 2006.
- [79] J. W. Creswell, *Qualitative Inquiry Research Design*, SAGE Publications, Inc, 3rd edition, 2013.
- [80] J. Pallant, *SPSS Survival Manual—a Step by Step Guide to Data Analysis using SPSS for Windows (Version 10)*, Routledge, London, 7th edition, 2020.
- [81] T. Sato, "An investigation of factors involved in Japanese students' English learning behavior during test preparation," *Papers in Language Testing and Assessment*, vol. 8, no. 1, pp. 69–95, 2019.
- [82] Y.-C. Pan, "Traditional and non-traditional test preparation practices: learner performance and perspectives," *Electronic Journal of Foreign Language Teaching*, vol. 3, no. 2, pp. 170–183, 2016.
- [83] D. S. Moore, W. I. Notz, and M. A. Fligner, *The Basic Practice of Statistics*, Freeman and Company, New York, NY, 6th edition, 2013.
- [84] J. David, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall, New York, 2nd edition, 2000.