

# Washback Effect of General English Test of Ph.D. Entrance Exam on Science and Humanities Students: Perceptions and Practices<sup>1</sup>

Fatemeh Abbasian Boroojeni<sup>2</sup>, Mansoor Tavakoli<sup>3</sup>, & Hossein Vahid Dastjerdi<sup>4</sup>

Received: 18/06/2016

Accepted: 8/11/2016

## Abstract

Testing holds an important place in the Iranian test-oriented context where the entrance examinations have serious educational and occupational outcomes. This study aimed to explore the potential washback effects of the General English Language Test of the Ph.D. (GELTP) as part of the Ph.D. University Entrance Examination (PUEE) on the humanities and science students' perceptions and practices. The study also examined the students' language learning needs to examine the correspondence between the test-taker's language needs as well as the GELTP requirement and content. To this end, the data were collected through conducting interviews with 16 Ph.D. students and the administration of a questionnaire to 560 students. Results revealed minor differences in the GELTP washback effects across the science and humanities perceptions and practices. Moreover, the findings showed some mismatch between the test-takers' language needs and the test content that affected negatively the test washback effects. Findings can have implications for raising the education and assessment authorities' awareness to employ the students' feedbacks in designing and modifying high-stakes tests.

**Keywords:** Washback; GELTP; Students' Perceptions; Students' Practices; Disciplinary Variations

---

<sup>1</sup>Please cite this paper as follows:

Abbasian Boroojeni F., Tavakoli, M., & Vahid Dastjerdi, H. (2017). Washback effect of general English test of Ph.D. entrance exam on science and humanities students: Perceptions and practices. *Journal of Research in Applied Linguistics*, 8(1), 109-132.

<sup>2</sup>Corresponding author, English Department, University of Isfahan, Isfahan, Iran; abbasianf@gmail.com

<sup>3</sup>English Department, University of Isfahan, Isfahan, Iran; mr.tavakoli14@gmail.com

<sup>4</sup>English Department, University of Isfahan, Isfahan, Iran; h\_vahid@yahoo.com

## 1. Introduction

There is a general consensus that high-stakes tests can have strong washback effects on learning (Qi, 2005; Spratt, 2005). In test-oriented contexts such as Iran, the use of high-stakes tests as gatekeepers for admission into higher education is very popular. One of the Iranian high-stakes tests that is administered annually to allocate the sparse places to studying in the Ph.D. program at state universities is the Ph.D. University Entrance Examination (PUEE). As an important part of the PUEE, the general English language test for Ph.D. (GELTP) that is common to all Ph.D. test-takers in different fields of study measures the students' general English proficiency knowledge. The test consists of different sections (i.e., grammar, reading comprehension, and vocabulary items), and it has gone through different modifications since its first administration in 2012 (see Appendix D). Therefore, test-takers compete for gaining access to the Ph.D. program and for achieving greater scholastic and professional opportunities, partly through enhancing their reading skill and general English vocabulary repertoire

In 2012, the Iranian measurement authorities launched some testing reforms to the format of the former PUEE. The examination was turned from a localized essay-type test administered under the supervision of each state university to a nationwide semicentralized multiple-choice test controlled by the Ministry of Science, Research, and Technology assessment commission. Based on the assessment authorities (ISNA, 2012), the main aim for the employed modifications was to have a standardized test with uniform and objective scoring procedures in order to increase the number of the PUEE applicants.

Due to the inefficacy of the Iranian educational system in enhancing students' English proficiency and fulfilling their language needs (Hayati, 2008), students at the tertiary level are often intimidated and challenged by the requirement of acquiring a high level of English knowledge to achieve satisfactory results on the entrance examinations. Also, the students are expected to have acquired enough English knowledge to fulfill their diverse language needs. Whereas undergraduate students need to have a good command of the general English, graduate students have to master highly specific academic English to cater for their target needs (Mazdayasna & Tahririan, 2008). However, because of power relations between the Iranian students and educational authorities, the students are often treated as passive recipients of different policies. They are not given enough autonomy to express their reactions and voices against the key educational aspects of curriculum development and assessment design (Khany & Tarlani-Aliabadi, 2016).

The importance of scrutinizing test-takers' language needs stems from the importance of English for the students at the tertiary level. In addition, English is

regarded as the language of international scholarship and as the medium of research communication for nonnative academics around the world (Hyland, 2007). According to Saif (2006), there is a connection between the way a test is designed and the way students learn. Consequently, the more congruity between the tests and students' needs, the more positive washback effects are expected. Douglas (2001) also recommends testing the language for specific purposes (LSP), rather than testing the general language because "LSP tests derive their content from an analysis of specific language use situations of importance to the test-takers" (p. 173).

Washback refers to the influence of testing on teaching and learning (Bailey, 1996). To investigate the washback effects of different tests, a number of frameworks are suggested by various scholars. In Hughes' (1993) trichotomy of backwash model, a test can affect participants, processes, and products. By suggesting the washback hypotheses model, Alderson and Wall (1993) made it possible for the washback phenomenon to be studied empirically. Based on Alderson and Wall's (1993) proposed hypotheses, "a test will influence learning, a test will influence what learners learn, and a test will influence how learners learn" (p. 121). As a combination of the two aforementioned models, in Bailey's (1996) basic model of washback, two types of washback to learners and programs are identified. However, because washback theories do not explore students' feedbacks, suggestions, and objections to different assessment agendas, needs analysis can be employed to complement such studies (Saif, 2006). According to Hutchinson and Waters (1987), needs analysis should be the starting point of any curriculum planning and assessment design. Therefore, as the first step in devising high-stakes tests, needs analysis can be adopted through considering and responding to stakeholders' views (Brown 2008).

## 2. Literature Review

In contexts with centralized educational systems, high-stakes tests are often used as agents for educational change (Cheng, 1998, 2005; Shohamy, Donitsa-Schmidt, & Ferman, 1996). However, due to the complexity of targeting washback as a result of indirect and unpredictable effects of tests on teaching and learning (Andrews, Fullilove, & Wong, 2002), the introduced changes in assessment processes are often not as effective as test developer expected they would be (Qi, 2005; Spratt, 1999). In different washback studies, some researchers have conceptualized the notion of washback (Alderson & Wall, 1993; Bailey, 1996; Hughes, 1993; Watanabe, 2004), whereas some others have examined the role of the engineered washback by investigating the effects of the modification to some tests on the stakeholders' practices and attitudes. These scholars believe that the high-stakes tests with important consequences can be manipulated to induce desirable

changes to the curriculum (Andrews, 2004; Qi, 2005; Wall, 2000), teaching (Cheng, 2005; Spratt, 2005), and learning (Andrews et al., 2002; Cheng, 1998; Spratt, 1999).

Among various stakeholders who are affected by different tests, tests are believed to have a powerful influence on language learners (Alderson and Hamp-Lyons, 1996).

To begin with, Cheng (1998 & 2005) examined the effects of the modifications made to Hong Kong Certificate of Education in English (HKCEE) on Hong Kong secondary school leavers' perceptions and behaviors through the cross-comparison of the results before the changes in 1994 and after the implementation of the changes in 1996. She employed extensive use of teacher and student questionnaires as well as the interviews with different stakeholders. The findings revealed that as a result of the change, the students tried to cope with the new process-oriented assessment. She concluded that whereas the changes in the test had some impact on the students' perceptions, the impact on the students' strategy learning and motivation was not fundamental.

Similarly, the influence of the College Student English Proficiency Test (CSEPT) and the General English Proficiency Test (GEPT) in 17 tertiary institutions in Taiwan was studied by Pan and Newfields (2012). A large number of students in nine institutions without English graduation requirements (EGR) and eight schools with EGR were interviewed. The results indicated no significant differences across the two groups' activities. The findings showed that although GEPT and CSEPT enhanced the learners' motivation, they did not have strong washback effects on the students' learning.

Andrews et al. (2002) investigated the effects of modifications to a high-stakes public test by the introduction of a new oral examination. They conducted the simulated oral tests with three groups of candidates with the goal of improving students' learning. The results indicated that the introduction of the Use of English (UE) oral examination caused some improvement in the students' performance in the spoken English but the precise nature of the washback seemed to be different from student to student.

Investigating the effect of high school final tests, Damankesh and Babaii (2015) investigated the washback effects of the test on the students' test-taking and test-preparation strategy use. They claimed that the tests have the power to influence the learners' strategic behaviors and learning. The results revealed that whereas the use of some strategies exerted a negative influence on the students' learning, some of the employed strategies fostered the students' linguistic and mental abilities.

Although students have the highest stakes and they are most affected by the consequences of tests (Hayes, 2003), the exploration of washback effects of high-stakes tests on students has been underresearched. In spite of the scholars' attention

to the washback phenomenon during the last few decades, most researchers have focused on the effects of examinations on teaching and teachers and few studies have explored the possible washback effects on learners and learning. The scarcity of the washback studies addressing learners is one of the impetus for undertaking the study.

Moreover, as a high-stakes test, the GELTP is expected to exert some degrees of washback effects on the test-takers' attitudes and practices. Therefore, the current study attempts to explore the potential washback effects of the GELTP on the humanities and science students' perceptions and practices. The comparative side of the study is undertaken to investigate the possible similarities and differences in the test washback effects across the two groups' attitudes and activities. Moreover, the level of consistencies between the GELTP requirements and content and the science and humanities students' language needs in the Ph.D. program are explored to see whether the employed changes to the GELTP are sufficient or more justified and well-planned modifications to the test are required.

### **3. Method**

#### ***3.1 Participants***

For the interview phase of the study, 16 Ph.D. students selected through convenience sampling were interviewed. To ensure the elicitation of both groups of humanities and science students' views, purposeful sampling, rather than random sampling, was used. Eight humanities students were recruited from the Politics, the Humanities and Social Sciences, and Literature and Language Faculties. Also, eight science students were recruited from the Science, Engineering, and Computer Faculties.

For the survey phase of the study, stratified random sampling was used to ensure a balanced representation of both humanities and science students. The questionnaires were distributed among 600 Ph.D. students (289 science and 311 humanities). From among the completed surveys, 560 (278 science and 282 humanities) questionnaires were analysed, and the rest with a high number of missing items were discarded.

#### ***3.2 Instruments***

In the first phase of the study, a semistructured interview protocol was used (see Appendix A). For devising the interview protocol, the review of the washback studies (Cheng, 2005; Qi, 2005; Sadeghi, 2012) as well as the students' comments and reflections in the two initial baseline interview sessions were employed.

To elicit the opinions of a large and potentially diverse population, the student questionnaire was developed and used. The first draft of the questionnaire was made based on the input obtained from the interview data and the reviewed

literature (e.g., Cheng, 1998, 2005; Tsagari, 2006; Mazdayasna & Tahririan, 2008; Sadeghi, 2012). To facilitate the analysis of the questionnaire and to see whether any patterns emerged from the data, the items in the student questionnaire were categorized into two sections of the students' perceptions and the students' behaviors.

To ensure the validity and reliability of the instruments, different steps were taken. To begin with, pre and postvalidation by insiders (i.e., the researchers) and outsiders (i.e., two colleagues) were achieved before the drafting of the instruments. Moreover, the initial drafts of the instruments were compared with similar ones in other washback studies and the contribution of every item towards the research questions was confirmed by the reviewers. Different data sources (i.e., interview and questionnaire) were also triangulated and converging codes and themes emerged. In addition, the initial version of the questionnaire was piloted on a group of 38 Ph.D. students at the University of Isfahan and the test-takers' judgments on the piloted instruments were gained. The piloting stage helped to eliminate ambiguous and irrelevant words and items. To compute the reliability of the questionnaire, Cronbach's alpha was used and it was  $r = 0.78$ . The reliability of the coding system in the interview was achieved through intercoder consistency. Some part of the data were selected randomly and recoded by the researchers and an expert colleague.

The final draft of the questionnaire had 18 items in two sections and six categories, as well as a question inquiring the students' field of study (see Appendixes B and C). The first section was related to the students' perceptions (items 1-10). This section consisted of three categories (i.e., constraints, skill enhancement priorities, and language needs assessment). The second section (items 11-18) was concerned with the students' behaviors and learning activities. This section contained three categories (i.e., preferred learning methods, participation in different test preparation courses, and test preparation materials). Both sections of the questionnaire were designed based on a five-point Likert scale from 5 (*strongly agree*) to 1 (*strongly disagree*).

### **3.3 Procedure**

In the first phase of the study, 10 interview sessions were conducted with 10 humanities and 10 science students. The Ph.D. students residing at the University of Isfahan dormitory were contacted and the meetings were arranged. The two initial sessions were considered as the baseline study. The aim of conducting the baseline study was to explore the participants' reactions and attitudes to the GELTP washback effects and encourage them to explain the reasons for their responses. Consequently, the participants' comments were used as the first source of information for developing the final draft of the interview protocol. For each

session, either two humanities or two science students were recruited for approximately 40 min of audio-taped interviews conducted in Persian. The interview protocol was, first, devised in English and then translated into Persian. All the interview data were, then, transcribed and translated back into English. The transcripts were repeatedly read through and analyzed. The similar segments were identified and labeled with a common code. The related codes were, then, grouped together to arrive at common themes. The main themes that were helpful in answering the research questions were identified and categorized.

In the second phase of the study, the data were gathered through the student questionnaires that were e-mailed to the Ph.D. students at state universities around the country or were distributed at various faculties at the University of Isfahan. The science and humanities students at different faculties were asked to fill out the questionnaire. After the distribution, a brief explanation was provided. It took the participants approximately 10 min to complete the questionnaire.

To analyze the data gathered through the student questionnaires, both descriptive and inferential statistics were used. First, the frequency distribution and percentages for all the items in the questionnaire were calculated. Tables were used to report the percentages. In the second step, the Mann-Whitney  $U$  test was employed for revealing the possible discrepancies or similarities between the science and humanities students' perceptions and behaviors. In reporting the results, when the  $p$  value was significant, the effect size was also provided to show that the differences were meaningful. To interpret the meaning of the  $r$  scores, Cohen's (1988) classifications for effect sizes as small ( $r = 0.2$ ), medium ( $r = 0.5$ ), and large ( $r \geq 0.8$ ) were used.

#### 4. Results and Discussion

In order to address the research questions, the potential washback effects of the GELTP on the humanities and science students' perceptions and behaviors were explored, and the possible similarities and differences of the test washback effects across the two groups' attitudes and activities were investigated. Moreover, the level of the correspondences and divergences between the students' needs and the GELTP content and requirements were examined.

In analyzing both the interview and questionnaire data, similar themes and categories emerged. Therefore, to discuss the results of the study, the findings in the two phases complemented each other to provide a well-supported report for each theme and category. To this end, the interview results that were relevant to the discussed topics and could help explain the questionnaire results were selected and reported along with the survey data.

#### 4.1 Students' Perceptions

In this study, students' perceptions refer to the students' understandings and attitudes regarding the content of the GELTP and the constraints imposed by the test. It also deals with the students' attitudes and feedbacks regarding the degree of alignment between their L2 needs and the test content and requirement. This section consisted of 10 items that explored the washback effects of the GELTP on the students' perceptions concerning three categories of constraints, skill enhancement priorities, and language needs assessment. The percentages of the level of agreement and disagreement across the parallel items in the two groups are summarized in Table 1:

Table 1 *Summary of Descriptive Statistics of GELTP Washback Effects on Science and Humanities Students' Perceptions*

Items	Science			Humanities		
	A	D	N	A	D	N
1	78.8	15.7	5.5	83.8	13.6	2.6
2	71.3	25.7	3.0	81.4	17.5	1.2
3	85.9	8.6	5.5	84.6	12.6	2.7
4	77.3	7.7	15.0	73.6	19.4	7.0
5	72.6	17.7	8.0	68.3	25.4	5.1
6	68.4	14.7	16.9	67.8	23.9	8.2
7	60.3	31.4	8.3	54.0	42.6	3.4
8	45.0	37.7	5.2	40.2	47.7	2.3
9	52.8	36.7	10.2	49.0	46.4	3.8
10	84.8	12.0	3.1	89.2	9.6	1.2

Note. (A: Agree = Strongly Agree and Agree; D: Disagree = Strongly Disagree and Disagree; N: Neutral = Neither Agree Nor Disagree; Total Number: 560)

The first category (items 1-2) was designed to examine the students' views regarding the GELTP imposed pressure and the time constraints when they were preparing for the test.

As for the first item, the findings revealed the detrimental washback effects of the test on the two groups' attitudes. The results showed that the majority of the science and humanities students (78.8% and 83.8%, respectively) believed that test pressure was one of their big concerns. As for the time constraint, the majority of the science (71.3%) and the humanities (81.4%) respondents believed that the time limitation hindered their preparation activities (see Table 1).

The GELTP negative washback effects, as the result of the test pressure, were pointed out by some of the interviewees, as well. They believed that the low level of their English proficiency and the time limitation were major constraints that



increased their level of anxiety and decreased the level of motivation to study for the test. One of the students said:

- *I was aware of the importance of having a good level of English language proficiency in my Ph.D. entrance exam and my Ph.D. program. Unfortunately, the educational system was not very effective in increasing our English knowledge. We had to handle all the pressures of learning English ourselves. Time limitation in this stage was a big concern. Because I could not cope with such challenge, I had to ignore the GELTP preparation altogether. I think the pressure imposed by the GELTP could enhance the students' efforts to learn English if the test and time pressures did not exceed the students' threshold level of tolerance.*

Almost all of the interviewed students complained about the lack of congruity between the discipline-specific needs and the GELTP content as a source of intensifying the test pressure and time constraints. One of them asserted:

- *With the Ph.D. test-takers' hectic schedule, we had to study general English and memorize lots of irrelevant vocabulary items to be prepared for the test. At least, if the test measured specialized English for each discipline, the time spent on the test preparation activities had more beneficial effects, and the level of motivation to cope with the test and time pressures would increase.*

The cross-comparison of the results (see Table 3) showed no significant differences in the test washback effects across both science and humanities students' perceptions ( $U = .066, p > .05$ ), revealing the concerns of both groups regarding the test pressures and the time limitation when preparing for the GELTP.

#### **4.2 Skill Enhancement Priorities**

This category (items 3-6) was designed to investigate the effects of the GELTP on the students' views regarding their skill priorities by placing specific emphasis on the skills which are needed most during the Ph.D. program. The test-takers' beliefs regarding the consistencies between the test content and their language needs can justify some of the students' test preparation activities.

As shown in Table 1, the importance of the reading skill (item 3) was unanimously acknowledged by both science and humanities students (85.9% and 84.6%, respectively) corresponding to Mazdayasna and Tahririan's (2008) findings that reading skill and improving the vocabulary repertoire were the most important skill and subskill priorities for the Iranian students. Therefore, the GELTP exerted a positive washback effect on the students' test preparation attitudes and activities because the reading skill was the main skill measured in the test.

The second most stressed skill for both groups was the writing skill (item 4). Interestingly, 77.3% of the science students and 73.6% of the humanities students believed in the necessity of the writing skill in their Ph.D. program (see Table 1).

According to the interviewed respondents, one possible justification for such strong need for the writing skill can stem from the requirement of publishing articles preferably in well-reputed journals as part of the Ph.D. graduation process in Iran. Despite the presence of some grammatical items in the GELTP, the majority of the interviewed students believed that because the writing skill was not assessed in the GELTP and due to the time constraints, they did not devote much time to its improvement in the test preparation activities. One of the interviewed students said:

- *The main reason for my reluctance to improve the writing skill is the low level of competence in this skill. All we know are some isolated grammatical structures that are assessed in different tests. Due to the time constraint and because the writing skill is not measured in the GELTP, I did not devote much time and effort to its learning. Sadly, the expectations from the Ph.D. students to publish articles are too high and challenging.*

For items 5 and 6, concerning the importance of speaking and listening skills for the Ph.D. students, the respondents' views were basically consistent. The science and humanities students' agreement on the necessity of speaking skill (72.6% and 68.3%, respectively) revealed the importance of this skill in the Ph.D. programs. Similarly, 68.4% of the science students and 67.8% of the humanities students asserted the need for the listening skill (see Table 1).

Although a large number of the interviewed students believed that learning these skills was very time-consuming, they complained about the lack of congruity between their target needs and the test content. One of the respondents said:

- *I strongly believe in the importance of the speaking and listening skills. In our Ph.D. program, we usually have presentation and seminar courses presented in English. Moreover, for taking part in the international conferences mastering these skills is necessary. Maybe, if they were measured in the GELTP, I would have devoted more time to their improvement.*

Based on Table 3, the cross-comparison of the results showed minor discrepancies ( $U = .000, p > .05$ ) across the two groups' perceived skills priorities. The effect size for this analysis ( $r = 0.09$ ) also indicated small differences between the two groups' skill priorities.

### 4.3 Language Needs Assessment

Apart from the limited number of items that measure isolated grammatical structures, the GELTP is mainly devoted to the assessment of the general English reading skill and vocabulary items (see Appendix D). Consequently, language needs assessment category (items 7-10) asked the respondents' views about the necessity of having other skills of writing, speaking, and listening as well as having the specialized English assessed in the GELTP.

As illustrated in Table 1, for items 7, 8, and 9, the percentages of the science and humanities respondents who agreed on having the skills of writing (60.3%, 54.0%), speaking (45.0%, 49.0%), and listening (52.8%, 49.0%) measured in the GELTP were almost the same. Therefore, the observed discrepancies between the science and humanities students (see Table 3) may be related to the differences in the level of the disagreement on having the skills assessed across the two groups due to the disciplinary-based needs and skill priorities in the Ph.D. program.

The interview results showed that the needs of the science students to writing, speaking, and listening skills and having them assessed were slightly higher than the humanities students' needs to these skills. The interviewed science students more frequently acknowledged the need to use scientific and English reference books and articles and write articles in well-reputed international journals as well as taking part in international conferences to have a more successful communication with the scientific world. On the other hand, some of the interviewed humanities students majoring in Arabic Language and Persian Literature disagreed with having other skills assessed in the GELTP because of their disciplinary requirements.

In agreement with Salehi and Yunus' (2012) results, that when a skill is measured in a high-stakes test, it is studied by the students, the interviewed students' perceived negative effects of the GELTP on the test preparation practices was partly due to the lack of measurement of different skills. They complained about the lack of having a skill-based test battery in which both productive (writing and speaking) and receptive (reading and listening) skills were assessed. One of them complained about this issue as follows:

- *I think some forms of assessment for different skills should be added to the GELTP. I know it sounds unpractical and it is difficult to assess such a large number of test-takers on listening and writing abilities. However, if these skills were assessed the same as the international tests such as TOEFL and IELTS, we might have placed more time and attention to these skills in our test preparation activities.*

Item 10 examined the students' views regarding having a specialized part related to each field of study in the PUEE for assessing the reading skill. A vast

majority of science and humanities students (84.8% and 89.2%, respectively) acknowledged the necessity of having a specialized English reading test rather than a general English reading test (see Table 1). The results were in agreement with Douglas' (2001) view that testing specialized language is preferred because each group has to control the technical language that is used in their professional or academic fields. The interviewed respondents' criticism over the absence of a specialized English test in the PUEE was also overwhelming. One of the students disapproving the one-for-all GELTP mentioned that:

- *There needs to be a specialized English test in the PUEE the same as what we had in our M.Sc. entrance examination. I do not see any justifications for moving backward to have the general English language test for the PUEE. In fact, in the Ph.D. program, discipline-specific English is, even more, important and our time is, even more, limited. However, although the GELTP has been implemented for five rounds and some modifications have been taken place, no response has been provided to the students' feedbacks and concerns about the inclusion of a specialized language test by the assessment authorities.*

The cross-comparison of the results in Table 3 indicated slight differences in the science and humanities students' perceptions on the necessity of having different skills assessed in the GELTP ( $U = .005, p > .05$ ). The effect size for this category ( $r = 0.08$ ) was also found to be very small.

#### 4.4 Students' Behaviors

The students' behaviors refer to the students' decisions and actions to prepare for the GELTP through self-learning activities, taking part in different preparation courses and the use of textbooks. This section consisted of 8 items (10-18) and three categories (i.e., preferred learning activities, participation in the test preparation courses, and selection of test preparation materials). As illustrated in Table 2, the summary of the results of the science and humanities students' behaviors indicated only minor differences in the percentages of the level of agreement and disagreement across the parallel items in the two groups.

Table 2 *Summary of Descriptive Statistics of GELTP Washback Effects on Humanities and Science Students' Practices*

Items	Science			Humanities		
	A	D	N	A	D	N
11	70.4	26.5	3.0	67.9	31.0	1.2
12	77.9	16.5	5.6	74.1	23.2	2.7
13	69.4	22.4	3.0	72.9	23.6	3.5
14	29.8	64.6	7.2	35.2	62.3	5.5
15	33.5	58.1	8.4	47.2	49.5	3.4

16	60.7	26.5	12.9	61.8	32.1	6.1
17	36.2	55.4	8.4	49.0	47.5	3.6
18	48.6	34.2	17.1	56.3	34.1	8.6

*Note.* (A: Agree = Strongly Agree and Agree; D: Disagree = Strongly Disagree and Disagree; N: Neutral = Neither Agree Nor Disagree; Total Number: 560)

#### 4.5 Preferred Learning Activities

This category was designed to explore the GELTP washback effects on the students' self-regulated learning activities through extensive reading and expanding the vocabulary repertoire.

Item 11 inquired about the students' self-regulated activities in English language learning. As shown in Table 2, a vast majority of the science and the humanities students (70.4% and 67.9%, respectively) believed in the importance of self-learning activities as part of a long-term test preparation process.

Item 12 asked about the students' preferred activities for the preparation of the GELTP by improving the reading skill through extensive reading practices and vocabulary improvement. As summarized in Table 2, a large number of the science and the humanities students (77.9% and 74.1%, respectively) believed in the importance of the reading skill practices.

The observed results, in line with Bailey (1996) that views the enhancement of the learners' autonomy as a sign for the beneficial effect of a test, revealed the positive effects of the GELTP on the students' self-learning and self-regulated reading-based activities. Because the main skill measured in the GELTP is reading, the students' emphasis on the reading skill improvement was justifiable.

All the interviewees unanimously emphasized the importance of self-regulated English learning activities. They asserted the importance of the English assessment at the tertiary levels as a strong stimulus for the students' self-learning activities from the early stages. This is in line with Wu's (2012) findings that the implementation of the GEPT in Taiwan enhanced the students' English learning through self-regulated activities. One of the respondents acknowledged this:

- *Compared to preparation activities for other entrance examinations, for the Ph.D. program, age-related factor and hectic schedule hinder our English learning activities. Because I was aware of the high demands of the Ph.D. program and of having a good command of English, I started learning English gradually at private institutions in my spare time to prepare myself.*

As shown in Table 3, the cross-comparison of the results revealed minor discrepancies across the two groups' preferred activities to increase their English

proficiency ( $U = .026, p > .05$ ). The obtained effect size ( $r = 0.06$ ) was also found to be trivial.

#### **4.6 Participation in Test Preparation Courses**

This category was designed to examine the effects of the GELTP on the students' activities through participation in different preparation courses. It consisted of three items (13-15) each considering a different preparation course.

Item 13 asked about the students' preferences in taking part in the general English proficiency courses to gain English knowledge. As illustrated in Table 2, the results revealed that both science and humanities students asserted the importance of taking part in English proficiency courses (69.4% and 72.9%, respectively).

The interviewed students homogeneously favored taking part in the general English proficiency courses as the best way to attain English knowledge as part of having a self-regulated English learning program. They blamed the inadequacy of the educational system in fulfilling their language needs as effectively as the private language institutions. Some of them mentioned that such courses could not be used as a quick test preparation remedy due to the time constraints. However, the importance of such courses for catering the students' academic needs in the Ph.D. program was asserted by the majority of the respondents. One of the students commented as follows:

- *Taking part in the English proficiency courses in private English institutions gradually and continually is the best way to learn English for our long-term academic needs. Nevertheless, it is more productive if the academic curriculum developers design effective English courses such as those run by private institutions for all the students around the country.*

Item 14 investigated the washback effect of the test in increasing the participants' tendency to take part in the preparation courses specially designed for the GELTP. The results in this part appeared to be paradoxical. On the one hand, the emphasis on the reading skill and the test-taking strategies were consistent with the GELTP requirements. On the other hand, the science and humanities students' perceived effectiveness of the GELTP preparation were very low (29.9 % and 39.9%, respectively; see Table 2).

The majority of the respondents in the interviews acknowledged the time shortage as the main justification for the lack of interest in taking part in the GELTP preparation courses. One of them stated that:

- *To attend test-preparation courses is a contradictory issue. While learning the test-taking strategies in the GELTP preparation courses can save us*

*lots of time and efforts through increasing our speed in handling the test, our time is too limited to be devoted to the preparation courses.*

Item 15 inquired the effect of the GELTP on the students' practices in taking part in other international test preparation courses such as IELTS or TOEFL as part of the GELTP preparation activities. The results revealed that the science and humanities students' interests in taking part in these courses were, to some extent, higher (33.5% and 47.2%, respectively) than participation in the GELTP preparation classes (see Table 2).

Based on the interviewed students' views, the importance of practicing the writing skills besides the reading skill motivated them to take part in the international test preparation courses. Moreover, they believed that the use of authentic exam-related materials in the international test preparation courses in comparison to the use of exam-related local materials in the GELTP preparation courses made the former more effective. For the majority of the students, the difficulty level of the GELTP and the slim chance of success in the test decreased their motivation to take part in the preparation courses. One of the students complained about this issue:

- *Due to the difficulty level of the test, I was reluctant to take part in the test preparation courses as I could not anticipate good results in the test. Even when I am asked for advice, I do not recommend preparation courses because of the limited results.*

As shown in Table 3, the cross-comparison of the results of the humanities and science students' practices revealed some discrepancies in the preparation activities for the GELTP between the two groups ( $U = .000$ ,  $p > .05$ ). For this category, the estimated effect size ( $r = 0.1$ ) was small. One possible explanation for the observed differences can be the higher agreement rate of the humanities students in participating in different preparation courses (see Table 2).

#### **4.7 Selection of Test Preparation Materials**

This category was designed to explore the GELTP washback effect on the students' preferences for selecting the exam-related materials. According to Read and Hayes (2003), test preparation materials range from highly exam technique oriented local materials to those authentic materials that attempted to improve the test-takers' English language abilities.

Whereas item 16 asked about the respondents' test preparation activities regarding the selection of the authentic test-preparation materials which are devised for international standardized tests (e.g., exam-related materials for IELTS or TOEFL), item 17 asked about the selection of the local materials prepared and

assembled by the Iranian authors. As it is summarized in Table 2, the results revealed a significant difference between the science and humanities students' preferences for the authentic materials (60.7% and 61.8%, respectively) in comparison with their preferences for the local test materials (36.2% and 49.0%, respectively). The science and humanities students' perceived inefficacy of the local test materials in preparing them for the test was manifested in the high percentages of their disagreement (55.4% and 49.0%, respectively) in choosing the local test-related materials as compared with the science and humanities students' disagreement levels over the selection of the authentic materials (26.5% and 32.1%, respectively).

Comparing the advantages and disadvantages of different exam-related materials in the interviews, the majority of the respondents favored the use of authentic materials. One of the students noted that:

- *It is easier for the students with the low level of English proficiency to follow the reading passages and vocabulary practices in the local exam-related textbooks because of the Persian translations and explanations. However, those local materials that are created through the modification of some authentic exam-related textbooks are not clearly organized. On the other hand, the exam-related authentic materials are more dependable because they put emphasis on enhancing all four skills concurrently.*

Item 18 asked about the respondents' use of the test papers and mock exams for promoting the test-taking skills and strategies and in familiarizing the test-takers with the sample questions of the test. Whereas 56.3% of the humanities and 48.6% of the science students acknowledged the effectiveness of using the test papers, a smaller number of the humanities and science students (34.1% and 34.2%, respectively) disagreed with the use of the previous test papers.

The comparative analysis of the results showed no significant differences in the GELTP washback effect on the science and humanities students' selection of the test-related materials ( $U = .183, p > .05$ ; see Table 3):

Table 3 *Mann-Whitney U Test Results: Comparison of Categories Across Humanities and Science Students*

	C1	C2	C3	C 4	C 5	C6
Mann-Whitney	52289.000	278931.500	115853.500	109242.000	244417.000	282288.000
Asymp. Sig. (2-tailed)	.066	.000*	.005*	.026*	.000*	.183

Note. C stands for categories,  $p > .05$

## 5. Conclusion and Implications

On the whole, the results in line with Loumbourdi's (2013) findings revealed a range of test washback from positive to neutral and to negative effects on



the students' perceptions and behaviors supporting the claim about the complexity of the washback phenomenon (Alderson & Hamp-Lyons, 1996; Cheng, 2005; Spratt, 2005). Moreover, the cross-comparison of the results revealed a high level of conformities in both negative and positive washback effects of the GELTP on the science and humanities students' perceptions and behaviors.

Concerning the students' attitudes, the results did not reveal significant discrepancies in the washback effects of the test across the two groups. The negative effects of the test on the students' perceived test pressures could stem from lowering the students' motivation due to the increased anxiety caused by the test constraints and time limitation. Such a result was in agreement with Cheng et al.'s (2014) views considering test anxiety and motivation as mutable interrelated factors. The cross-comparison of the results in the categories "skill enhancement priorities" and "language needs assessment" revealed minor differences in both the facilitating and debilitating washback effects of the GELTP on the science and humanities students' perceptions. The discrepancies can be related to the students' diverse language learning needs because of the disciplinary-based requirements. The test exerted positive washback effects on the students' attitudes to improve the reading skill that was assessed in the test. However, because all the Ph.D. test-takers in different fields of study have a common GELTP, the test imposed some negative effects on the students' attitudes. The majority of the interviewed and surveyed students acknowledged the need for having disciplinary-based specialized English assessed in the PUEE. Moreover, the test exerted some negative washback effects on the test-takers' perceptions through decreasing their motivation and effort to improve the writing, speaking, and listening skills that were not measured in the test.

As regards the science and humanities students' practices, the findings showed some differences across the two groups. It seemed that the content and requirement of the GELTP influenced the students' activities by directing them towards preparing for the test. The findings of the category "preferred learning activities" showed small discrepancies across the science and humanities students' perceived importance of the role of self-learning in succeeding in the test and fulfilling the long-term needs of enhancing the English ability. Consequently, in agreement with Bailey (1996) that views the enhancement of the learners' autonomy as a sign for the beneficial effect of a test, the effect of the GELTP on the students' self-learning and improvement of the reading skill was positive. The motivating role of the English assessment in the entrance examinations at the tertiary levels for enhancing the students' English learning through self-regulated activities is in accordance with Wu's (2012) views in the implementation of the GEPT in Taiwan. Furthermore, the results indicated small differences in the washback effects of the test on the science and humanities activities regarding the category "participation in

the test preparation courses.” The differences can be explained through the humanities students’ higher interest in taking part in different test preparation courses. Nevertheless, one possible justification for the students’ negligible interest in participation in different English preparation courses was the time constraints. Moreover, based on the interview results and in line with Xie and Andrew (2012), the students’ limited interest in the GELTP preparation activities may have been initiated from the test-takers’ negative perceptions of the difficulty level of the test and, consequently, the low anticipation of the test success. Concerning the last category “selection of test preparation materials,” the test exerted positive washback effects on the two groups. Both groups showed high preferences for using authentic test-preparation materials. The results also indicated the test neutral washback effect on the test-takers’ activities for the use of past papers and mock test.

In general, the GELTP exerted positive effects on the test-takers’ perceptions and activities to improve the reading skill that was measured by the test and was one of the students’ basic academic priorities. However, the results indicated the test-takers’ discontent for not having the specialized disciplinary-based English and other skills (i.e., writing, listening, and speaking) assessed in the PUEE. Despite the test designers’ continuous modifications to the GELTP during the last five rounds of the test administration since 2012, no attention has been paid to the test-takers’ suggestions and concerns. The problem may partly arise from the assessment authorities’ disregard for the implementation of needs analysis prior to devising or modifying large-scale high-stakes tests. In line with Khany and Tarlani-Aliabadi (2016), students are basically viewed as the passive and powerless recipients of predetermined policies in Iranian educational system. To bridge the gap, informed interactions is necessary between the assessment authorities and the recipient of the programs and policies. According to the interviewed respondents, the students have never been given any chances to voice their concerns. They believed that the assessment authorities’ fixed policies should be changed into more flexible practices through engaging in ongoing reappraisal and readjustment of the test development processes by taking the test-takers’ feedbacks into account.

The study can have practical implications for broadening the assessment authorities’ and test developers’ perspectives regarding the evaluation and the test-takers’ language needs. Through raising the assessment authorities’ awareness and informing them about the solicited students’ views, they can align the PUEE assessment policies purposefully with the students’ language needs.

Finally, the absence of the data elicited from other stakeholders such as the assessment authorities and the test designers is one of the limitations of the study. Unfortunately, in the Iranian context, due to the power relation issues and the educational authorities’ reluctance to implement students’ views in decision-making

processes, it is a highly unlikely endeavor to be connected to the assessment authorities and solicit their perspectives and feedbacks.

### References

- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13(3), 280-297.
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129.
- Andrews, S., Fullilove, J., & Wong, Y. (2002). Targeting washback-a case-study. *System*, 30(2), 207-223.
- Andrews, S. (2004). Washback and curriculum innovation. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp.37-50). Mahwah, New Jersey: Lawrence Erlbaum.
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257-279.
- Brown, J. D. (2008). Testing-context analysis: Assessment is just another part of language curriculum development. *Language Assessment Quarterly*, 5(4), 275-312.
- Cheng, L. (1998). Impact of a public English examination change on students' perceptions and attitudes toward their English learning. *Studies in Educational Evaluation*, 24(3), 279-301.
- Cheng, L. Y. (2005). *Changing language teaching through language testing: A washback study*. Cambridge: Cambridge University Press.
- Cheng, L., Klinger, D., Fox, J., Doe, C., Jin, Y., & Wu, J. (2014). Motivation and test anxiety in test performance across three testing contexts: The CAEL, CET, and GEPT. *TESOL Quarterly*, 48(2), 300-330.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Earlbaum Associates.
- Damankesh, M., & Babaii, E. (2015). The washback effect of Iranian high school final examinations on students' test-taking and test-preparation strategies. *Studies in Educational Evaluation*, 45, 62-69.
- Douglas, D. (2001). Language for Specific Purposes assessment criteria: where do they come from? *Language Testing*, 18(2), 171-185.
- Hayati, A. M. (2008). Teaching English for special purposes in Iran: Problems and suggestions. *Arts and Humanities in Higher Education*, 7(2), 149-164.

- Hayes, B. M. (2003). *IELTS preparation in New Zealand: An investigation into the nature of the courses and evidence of washback*. Unpublished doctoral dissertation, Victoria University of Wellington, New Zealand.
- Hughes, A. (1993). *Backwash and TOEFL 2000*. Unpublished manuscript. University of Reading, Malaysia.
- Hutchinson, T., & Waters, A. (1987). *English for specific purposes: A learning-centered approach*. Cambridge: Cambridge University Press.
- Hyland, K. (2007). English for professional academic purposes: Writing for scholarly publication. In D. Belcher (Ed.), *Teaching language purposefully: English for specific purposes in theory and practice* (pp. 20-45). New York: Cambridge University Press.
- Khany, R., & Tarlani-Aliabadi, H. (2016). Studying power relations in an academic setting: Teachers' and students' perceptions of EAP classes in Iran. *Journal of English for Academic Purposes, 21*, 72-85.
- ISNA, (2012, November 21). *Why did the Iranian Ph.D. program entrance exams altered to be semicentralized*. Retrieved April 20, 2016, from the World Wide Web: <http://isna.ir/fa/news/91090100296>
- Loumbourdi, L. (2013). *The power and impact of standardized test: Investigating the washback of language exams in Greece*. Internationaler Verlag der Wissenschaften: Peter Lang GmbH.
- Mazdayasna, G., & Tahririan, M. (2008). Developing a profile of the ESP needs of Iranian students: The case of students of nursing and midwifery. *Journal of English for Academic Purposes, 7*(4), 277-289.
- Pan, Y.-C., & Newfields, T. (2012). Tertiary EFL proficiency graduation requirements in Taiwan: A study of washback on learning. *Electronic Journal of Foreign Language Teaching, 9*(1), 108-122.
- Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing, 22*(2), 142-173.
- Read, J., & Hayes, B. (2003). The impact of IELTS on preparation for academic study in New Zealand. *IELTS Research Reports, 4*, 154-205.
- Sadeghi, S. (2012). *A critical reflection on the high-stakes test' effects: Towards a local model for teacher professionalism in Iranian preparation courses*. Unpublished doctoral dissertation, University of Isfahan, Isfahan, Iran.
- Saif, S. (2006). Aiming for positive washback: A case study of international teaching assistants. *Language Testing, 23*(1), 1-34.

- Salehi, H., & Yunus, M. M. (2012). The washback effect of the Iranian universities entrance exam: Teachers' insights. *GEMA Online® Journal of Language Studies*, 12 (2), 609-628.
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13(3), 298-317.
- Spratt, M. (1999). How good are we at knowing what learners like? *System*, 27(2), 141-155.
- Spratt, M. (2005). Washback and the classroom: The implications for teaching and learning of studies of washback from exams. *Language Teaching Research*, 9(1), 5-29.
- Tsagari, K. (2006). *Investigating the washback effect of a high-stakes EFL exam in the Greek context: Participants' perceptions, material design and classroom applications*. Unpublished doctoral dissertation, Lancaster University, England.
- Wall, D. (2000). The impact of high-stakes testing on teaching and learning: Can this be predicted or controlled? *System*, 28(4), 499-509.
- Watanabe, Y. (2004). Methodology in washback studies. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp.19-36). Mahwah, New Jersey: Lawrence Erlbaum.
- Wu, J. R. (2012). GEPT and English language teaching and testing in Taiwan. *Language Assessment Quarterly*, 9(1), 11-25.
- Xie, Q., & Andrews, S. (2012). Do test design and uses influence test preparation? Testing a model of washback with Structural Equation Modeling. *Language Testing*, 30(1), 49-70.

## **Appendix A**

### **Interview Protocol**

#### **A. Students' perceptions with regard to the GELTP**

1. Was English one of your main concerns when you decided to take part in the Ph.D. exam? Why?
2. How much time did you spend on the GELTP preparation? Was time a big concern?
3. Which of the reading, writing, speaking, and listening skills are needed the most in the Ph.D. program? Why?
4. On which skill (reading, writing, speaking, and writing) do you place more emphasis when you are preparing for the GELTP? Why?
5. Do you recommend having other skills assessed besides the reading skill? Which skills? Why?

**B. Students’ behaviors with regard to the GELTP**

6. Has the implementation of the GELTP led to your using different self-learning a? (e.g., reading texts extensively, memorising a large number of vocabularies, and studying grammatical rules)
7. Was taking part in different preparation classes your preferred activity when you were preparing for the GELTP? Why?
8. Which kind of preparation classes did you prefer (e.g., general proficiency courses, the GELTP courses or preparation classes for other high-stakes tests)?
9. When preparing for the GELTP, did you prefer using authentic or local materials? Why?
10. When preparing for the GELTP, did you prefer using the past test papers? Why?

**Appendix B  
Student Questionnaire**

Dear students,

This study aims to investigate the washback effects of the general English language test of the Ph.D. exam (the GELTP) on your attitudes and activities with special focus on your short- and long-term language learning needs. Please fill in the questionnaire based on your own opinion. The provided responses will be confidential and will be used for the research purposes only. Thank you for your cooperation.

(Please specify your field of study .....)

Please grade the following on a 5-point scale format: 5 (*Strongly Agree*), 4 (*Agree*), 3 (*Neutral: Neither Agree Nor Disagree*), 2 (*Disagree*), 1 (*Strongly Disagree*).

Items	5	4	3	2	1
	(SA)	(A)	(N)	(D)	(SD)
<b>Section one: Students’ perceptions</b>					
1	When taking part in the Ph.D. exam, the GELTP is one of the main concern.				
2	To prepare for the GELTP, the time constraint is one of the test-takers' concern.				
3	Reading is the skill needed in the Ph.D. program.				
4	Writing is the skill needed during the Ph.D. program.				
5	Speaking is the skill needed during the Ph.D. program.				
6	Listening is the skill needed during the Ph.D. program.				
7	It is necessary to have the test-takers' writing skill assessed in the GELTP.				

- 
- 8 It is necessary to have the test-takers' speaking skill assessed in the GELTP.
- 9 It is necessary to have the test-takers' listening skill assessed in the GELTP.
- 10 It is necessary to have the test-takers' reading skill assessed through a disciplinary-based specialised English knowledge.
- Section two: Students' behaviors**
- 11 Your English knowledge needed for the GELTP is gained through the self-regulated activities.
- 12 For a test-taker, the best way to prepare for the GELTP is to learn vocabulary and to read extensively.
- 13 Your English knowledge needed for the GELTP is gained through taking part in the General English proficiency Courses
- 14 Your English knowledge needed for the GELTP is gained through taking part in the GELTP preparation courses.
- 15 Your English knowledge needed for the GELTP is gained through taking part in other high-stakes tests preparation courses such as IELTS.
- 16 The best way to prepare for the GELTP is to work with authentic test-related materials prepared for the high-stakes test such as TOEFL and IELTS.
- 17 The best way to prepare for the GELTP is to work with the local material.
- 18 To prepare for the GELTP doing previous test papers and sample questions are helpful.
- 

### Appendix C

#### Summary of Items and Categories of the Questionnaire

Items	Categories	Section one: Students' perceptions
1	1. Constraints	Test pressure
2		Time constraint
3		Reading as target skill needs
4	2. Skill enhancement priorities	Writing as target skill needs
5		Speaking as target skill needs
6		Listening as target skill needs
7	3. Language needs assessment	Necessity of having writing skill assessed
8		Necessity of having the speaking skill

---

9		assessed Necessity of having the listening skill
10		assessed Necessity of having specialised English
11	4. Preferred learning activities	<b>Section two: Students' behaviors</b> Self-learning
12		Extensive reading and vocabulary improvement
13	5. Participation in different test preparation courses	Taking part in the General English proficiency Courses
14		Taking part in the GELTP preparation courses
15		Taking part in other high-stakes tests preparation courses
16	6. Selection of test preparation materials	The use of authentic test preparation materials
17		The use of local test preparation materials
18		The use of test prepares

**Appendix D**  
**Components of GELTP**

The date of the test administration	2012	2013	2014	2015	2016
The test content					
Grammar	20	20	8	8	8
Vocabulary	25	20	12	12	12
Cloze test	15	10	0	0	0
Reading comprehension	40	20	10	10	10
Total number of items	100	70	30	30	30

*Note.* This table is based on the test samples from 2012-2016, retrieved from the <http://Ph.D.test.ir>