



An Investigation into Iranian Non-English PhD Students' Perceptions Regarding Learning as an Educational Consequence of EPT

Mahbubeh Rezaeian

PhD. Candidate, Department of English Language Teaching, Aliabad Katoul Branch, Islamic Azad University, Aliabad Katoul, Iran

Seyyed Hassan Seyyedrezaei, Ph.D.

Department of English Language Teaching, Aliabad Katoul Branch, Islamic Azad University, Aliabad Katoul, Iran

Ghasem Barani, Ph.D.

Department of English Language Teaching, Aliabad Katoul Branch, Islamic Azad University, Aliabad Katoul, Iran

Zari Sadat Seyyedrezaei, Ph.D.

Department of English Language Teaching, Aliabad Katoul Branch, Islamic Azad University, Aliabad Katoul, Iran

Abstract

It appears that English Proficiency Test (EPT) as a high-stakes test plays a key role in addressing many aspects of the educational system and individuals' lives. This paper aimed to represent Iranian non-English PDD students' perceptions regarding their learning and testing preferences, and to investigate the relationship between educational consequences and psychological consequences of EPT among Iranian non-English PhD students. To reach the goals, two Persian researcher-made questionnaires entitled "educational consequences questionnaire" and "psychological consequences questionnaire" constructed and validated by Rezaeian, Seyyedrezaei, Barani, and Seyyedrezaei, (2020) were utilized. To ensure reliability, a pilot study was conducted on 60 participants and the content validity of the questionnaires confirmed by four experts in the field of language testing and assessment; afterwards, the questionnaires were distributed among 252 students via online administration throughout Iran. The analyzed results by the latest version of SPSS showed that deep learning had the highest mean among other subscales and item 3 in testing preferences which was related to creativity hit the lowest mean score. Moreover, the results confirmed that there was a significant and positive relationship between educational consequences and psychological consequences.

Keywords: Educational consequences, learning, psychological consequences, testing preferences

Introduction[#]

As stated by Shohamy (2001b), tests are responsible for turning the pleasure and enjoyment of learning into pain, nervousness, and a feeling of inequality. Tests are frequently the source of irritation, frustration, rivalry, pressure, and humiliation (Shohamy, 2007b, p. 142). As Bachman (1990) affirms, because testing occurs in an educational and social context, it is necessary for researchers to investigate how language tests have an effect on individuals, teachers, societies, teaching, learning, students, schools, language, and language policies. It is the power of tests, especially high-stakes tests since they cause test-takers as well as

educational systems to alter their educational performances, behavior and strategies as they endeavor to be successful in tests (Shohamy, 2017). Additionally, it has been noted by Menken (2017) that tests manipulate massive power over the lives of learners as well as educators and can shape in what way testing policy is applied in schools and societies. Considering the significance of tests, testers have begun to give attention to the uses, influence and consequences of tests and their function not only in educational contexts, but also in social, political in addition to economic contexts in recent years (Shohamy, 2001b). Administrators make some important decisions based on the test-takers' general language proficiency. These decisions evaluate their general knowledge or skills requirement to enter or

[#]Corresponding Author

Email: srezaei.sh@gmail.com

Received: 07/02/2020

Accepted: 11/23/2020

exit from the universities and institutions (J. D. Brown, 2008) and tests are given more power than ever recently, because they are extensively used via governments, institutions, in addition to central authorities in the world so individuals are judged through their language proficiency (Shohamy, 2007b).

As far as the history is concerned, the knowledge of psychometrics derived from work on intelligence testing. The fundamental belief was that intelligence was innate and unchanging in the way that other innate characteristic traits as intelligence could be measured because it was visible similar to other characteristics and on the basis of the result people could be appointed to groups, streams or schools which were suitable to their intelligence (Gipps, 2012). H. D. Brown (2004) defines testing as a way of conducting assessment which is strictly associated with fixed timing and stable procedures. As mentioned by Gipps (2012), one of fundamental limitations of the traditional psychometric testing model is measuring characteristics which are a property of the human beings and which are seen to be fixed. Shohamy (2017) argues that while psychometricians have been successful to develop complicated methods for developing and designing tests considering reliability, validity in addition to quality of items and tasks, they are inclined to ignore the significant dimension of consequences of tests and have failed to pose and answer several questions pertinent to tests' purposes. The association between testing and learning is confirmed by washback (Bachman & Palmer, 1996). In fact, they have viewed washback as a test's impact on individuals, society and educational systems. According to them, this washback effect can be either beneficial or harmful. Traditional assessment has been criticized by a number of researchers. Namely, Law, and Eckes (2007) imply that traditional assessments are not neither informative about the improvement of a student nor effective in the difficulties and challenges that they may encounter while answering the test. Additionally, Genesee and Hamayan (1994) believe that these types of tests are valuable for collecting information about students' achievements under specific circumstances, but they are ineffective to provide information about students' nonlinguistic factors such as individual differences, motivations, interests and their learning strategies just to name a few. Furthermore, they encourage students to concentrate surface and rote approaches to learning rather than deep and meaningful learning as passive learners (Newstead & Findlay, 1997).

EPT is an example of such proficiency tests which is conducted by the Ministry of Science, Research, and Technology once in a month from 1386 up to now in a

variety of authorized centers across Iran. This exit test is required for the PhD candidates of Islamic Azad universities who have chosen to continue their studies in Iran. More specifically, EPT comprises three different parts including vocabulary (25 questions in multiple-choice format), grammar (40 questions in multiple-choice and error correction formats) and reading comprehension (35 questions in multiple-choice and cloze test formats) in which test-takers are given 140 minutes to answer the questions. However, reading, writing, and listening skills are not assessed in this test. The test-takers' acquired knowledge until the exam time will be evaluated and no consideration is paid to the way in which they have attained such knowledge or skills (Noori & Hosseini Zadeh, 2017). EPT is considered as a norm-referenced proficiency test (Noori & Hosseini Zadeh, 2017) which is scored objectively via the computerized objective scoring systems to estimate the general competence of the test-takers in comparison with other take-takers. The take-takers can observe their result sheets online after seven to ten days from the examination time. Additionally, there is no limitation for the exam registration and they can register for the upcoming examinations in the case that they fail to get the required minimum cut-off score which is 50 out of 100.

As argued by Dhindsa, Omar, and Waldrip (2007), investigating students' perceptions of assessment encourages students to build up an authentic and realistic assessment. They also mentioned that students should be responsible for their learning. Struyven, Doch., and Jassess (2005) revealed that students' perceptions toward assessment influenced deeply the way they learned and studied. Given the fact that assessment notably influences students' approach to learning, assessment paradigms have shifted in 20th century from "testing learning of students to assessing for students learning" (Birenbaum & Felman, 1999, p. 92). According to Poehner (2013), in formal testing situations, it is required for students to perform in isolation without any interaction; moreover, in this context, they do not receive any external support or scaffolding. A considerable body of literature has been published on washback. These studies have demonstrated that how deeply rooted in social and education systems the current language testing process is. As an example, a survey study in this area conducted by Al Amin and Greenwood (2018) showed that due to washback effects, teachers diminished classroom teaching and learning to a curriculum focused almost totally on what was expected in the test. Additionally, testing was likely to decrease academic progress, applied huge pressure on students' lives. Interestingly, the chance to earn further income

via after-school coaching also persuaded low-paid teachers to alter their attention from making effort to meet national curriculums objectives to focusing on the limited framework of examinable materials (Al Amin & Greenwood, 2018). A recent study by Castro and Vega (2017) confirmed that washback effect could change students' perceptions and attitudes toward language testing in various ways. A range of the collected data confirmed that students were influenced harmfully from high-stakes testing. For instance, they were anxious and frustrated about testing (Fitzgerald, 2015). According to Pan (2009b), some students endeavored greatly to study English regarding the exit requirement since they were afraid of not being able to graduate. In a fascinating study conducted by Wheelock, Bebell, and Haney (2000), students drew themselves as irritated, nervous, tired, and negative when asked to illustrate a self-portrait in testing situations.

By similarity, the findings of the recent study done by Mohammad Salehi and Tarjoman (2017) in Iran indicated that washback considerably influenced both the learning and teaching methodologies. They found that MA Entrance Exam divided the curriculum into relevant and irrelevant or important and unimportant sections since teachers got interested in improving the scores of the students; thus, it created fear in students in the process of learning. Furthermore, the results of their study specified that every university professors and the most of the students wanted to have control over some matters such as the content of the test and the time of the administration. Students also demanded that MA Entrance Exam should be applied as an indicator of their language ability or knowledge instead of test-taking skills. A correlational study revealed that test anxiety had debilitating impacts in language learning and it was negatively associated to foreign language test performance (Massomeh Salehi & Marefat, 2014). A more recent study by Rezaeian, Seyyedrezaei, and Seyyedrezaei, (in press) showed that Iranian non-English PhD students were intrinsically motivated to be prepared for EPT. In addition, they demonstrated the low level of self-efficacy beliefs towards their achievement in EPT. According to this study, a high amount of test anxiety, stress, hopeless, nervousness, families' stress or tension, amotivation in addition to university dropout rate were discovered among the participants. Spearman results also verified that there was a significant correlation between motivation and self-efficacy beliefs, motivation and feelings, in addition to self-efficacy beliefs and feelings. Some other studies conducted on washback throughout the world from washback studies from 2017 to 2020 are: Munoz

(2017); Khoshshima, Saed, and Mousaei (2018); Hung and Huang (2019); Dinh (2020). Besides, several studies have been published in the field of washback of high-stakes testing in Iranian context such as Estaji and Tajeddin (2012); Ostovar Namaghi (2013); Kheirkhah and Ghonsooly (2014); Khodabakhshzadeh, Zardkanloo, and Alipoor (2017) Siahpoosh, Ramak, and Javandel (2019) just to name a few.

To date, it seems that there is a relatively small body of literature on the subject of Iranian non-English PhD students' perceptions toward EPT. Accordingly, this paper seeks to critically examine what Iranian non-English PhD students' perceptions are toward learning as an educational consequence of EPT since students are active participants or agents, researchers are in charge of promoting the students' voice both in learning and assessment (Gilmore & Smith, 2008). Presumably, the perceptions of students in high-stakes testing have not been addressed in the published literature in much detail and their voices cannot be heard or even have been rejected instead there are several studies based on teachers' perceptions as main stakeholders such as Gunn, Al-Bataineh and Abu Al-Rub (2016) as well as Thibodeaux (2014). This paper also aims to represent Iranian non-English PhD students' perceptions regarding their testing preferences, and to investigate relationship between educational consequences and psychological consequences of EPT among Iranian non-English PhD students. In particular, this paper will examine the following main research questions:

1. What are Iranian non-English PhD students' perceptions toward learning as an educational consequence of EPT?
2. What are Iranian non-English PhD students' perceptions toward testing preferences?
3. Is there any significant relationship between educational consequences and psychological consequences of EPT among Iranian non-English PhD students?

In order to answer the above-research questions the following research hypothesis will be tested:

H0: There is no significant relationship between educational consequences and psychological consequences of EPT among Iranian non-English PhD students?

Method

Participants

A total of 252 Iranian non-English PhD students took part in this study based on convenience sampling.

They were both male (N=139) and female (N=113) from different Islamic Azad universities in Iran. Table

1 tæser'eg e aarticipatt s' distribution by age.

Table 1.

Distribution of Participants by Age

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	less than 30	16	6.3	6.3	6.3
	30-35	64	25.4	25.4	31.7
	36-40	76	30.2	30.2	61.9
	41-45	47	18.7	18.7	80.6
	46-50	40	15.9	15.9	96.4
	more than 50	9	3.6	3.6	100.0
	Total	252	100.0	100.0	

Instruments

In order to carry out the present research, two attitudinal researcher-made questionnaires entitled "educational consequences questionnaire" and "psychological consequences questionnaire" developed and validated (in Persian) by Rezaeian et al. (2020) were utilized at exploring educational consequences and psychological consequences of EPT among Iranian non- English PhD students. Furthermore, the face and content validity of these researcher-made questionnaires confirmed by four experts in the field of language testing and assessment.

Educational Consequences Questionnaire: ECQ consists of five subscales, 23 items at a five-point Likert-scale format scaling from strongly disagree to strongly agree (Appendix I). The reliability of the questionnaire was estimated to be 0.93 in the main study through Cronbach's alpha coefficients. Hence, the questionnaire was reliable enough with the sample of study.

Subscale 1: Learning Environment (questions 1- 4)

Subscale 2: Surface Learning (questions 5-6)

Subscale 3: Deep Learning (questions 7-16)

Subscale 4: Learning Outcomes (questions 17-19)

Subscale 5: Testing Preferences (questions 20-23)

Psychological Consequences Questionnaire: PCQ comprises four subscales with 15 items at a five-point Likert-scale format scaling from strongly disagree to strongly agree (Appendix II). The reliability of the scale computed through Cronbach's alpha was 0.89 with the sample study.

Subscale 1: Self-Efficacy Beliefs (questions 1- 3)

Subscale 2: Feelings (questions 4- 9)

Subscale 3: Intrinsic Motivation (questions 10- 13)

Subscale 4: Extrinsic Motivation (questions 14- 15)

Procedure

At the outset, as noted above, two attitudinal researcher-made questionnaires named "educational consequences questionnaire" and "psychological consequences questionnaire" developed and validated (in Persian) via reviewing the allied literature, analyzing four experts' opinions, documents, semi-structured interviews in addition to conducting a number of statistical procedures to achieve the purposes of the study (Rezaeian et al., 2020). Statistical procedures were conducted at Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) as two different phases. To make sure of the reliability, a pilot study was carried out on 60 samples who were similar to the target participants. Subsequently, a total number of 252 Iranian non-English PhD students from different Islamic Azad universities completed the final researcher-made questionnaires by way of online administration throughout Iran. To analyze the data and answer the descriptive as well as correlational research questions, SPSS, version 24 was used in the present study.

Findings

First of all, in order to ensure the normality of data, Kolmogorov-Smirnov (KS) Test and Shapiro-Wilk were used. As Table 2 presents, since α value of both KS Test $.04 < .05$ and Shapiro-Wilk $.01 < .05$ for ECQ and KS Test $.001 < .05$ as well as Shapiro-Wilk $.001 < .05$ for PCQ are less than $.05$; therefore, the distribution of the data is not normal and nonparametric tests will be utilized for further analysis.

Table 2.*Test of Normality*

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
PCQ	.077	252	.001	.979	252	.001
ECQ	.058	252	.041	.985	252	.011

a. Lilliefors Significance Correction

Note. PCQ= Psychological Consequences Questionnaire;

ECQ= Educational Consequences Questionnaire

Research Question One

What are Iranian non-English PhD students' perceptions toward learning as an educational consequence of EPT?

As indicated in Table 3, participants perceived deep learning higher ($M = 39.81$) than the other three constructs. The second place is taken by learning environment with mean score 13.03. Learning outcome and surface learning have occupied the third and fourth place with mean scores 9.30 and 5.86 respectively.

Table 3.*Descriptive Statistics of Educational Consequences*

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
LE	252	4.00	20.00	13.0397	3.86299	.264	.153	-.807	.306
SL	252	2.00	10.00	5.8690	2.29904	.396	.153	-.814	.306
DL	252	10.00	50.00	39.8135	9.45112	-.872	.153	.066	.306
LO	252	3.00	15.00	9.3095	3.79655	.330	.153	-1.242	.306
Valid N (listwise)	252								

Note. LE= learning Environment; SL= Surface Learning; DL= Deep Learning; LO= Learning outcome

Research Question Two

What are Iranian non-English PhD students' perceptions toward testing preferences?

As shown in Table 4, item 1(I prefer questions that require more or deeper thinking) hits the highest place

with mean score 3.54. Item 2 which is (I prefer all language skills, including speaking and writing, to be assessed in this test) and item 4 (I prefer questions that need further analysis) have the second and third mean scores 3.48 and 3.43 respectively. The table also displays that item 3 (I prefer questions that require creativity to be answered) has the lowest mean score ($M=3.31$).

Table 4.*Descriptive Statistics of Testing Preferences*

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Q1	252	1	5	3.48	1.349	-.414	.153	-1.121	.306
Q2	252	1	5	3.54	1.212	-.367	.153	-1.013	.306
Q3	252	1	5	3.31	1.323	-.109	.153	-1.322	.306
Q4	252	1	5	3.43	1.253	-.275	.153	-1.156	.306
Valid N (listwise)	252								

Research Question Three

Is there any significant relationship between educational consequences and psychological consequences of EPT among Iranian non-English PhD students?

Spearman correlation coefficient was used to determine the relationship between PCQ and ECQ in order to answer the third research question. The result of correlation analysis which is presented in Table 5 manifests that there is a strong and positive correlation between educational consequences and psychological consequences ($r = .774$, $n = 252$, $p = .000$).

Table 5.*Correlation between Educational Consequences and Psychological Consequences*

			PC	EC
Spearman's rho	PC	Correlation Coefficient	1.000	.774**
		Sig. (2-tailed)	.	.000
		N	252	252
	EC	Correlation Coefficient	.774**	1.000
		Sig. (2-tailed)	.000	.
		N	252	252

** . Correlation is significant at the 0.01 level (2-tailed).

Note. PC= Psychological Consequences; EC= Educational Consequences

Discussion and Conclusion

According to Kwako (2003), the strongest point of traditional assessments is the ease in designing and scoring. Consequently, they are time- effective and effortless to create appropriate tests. Most important of all is the issue objectivity in traditional assessments. Apart from strengths, traditional assessment suffers from some pitfalls as follows: It enhances competitive spirit for obtaining higher scores among a few top students; students encourage working alone and deemphasizing the collaboration as well as shared interaction among students (Helmericks as cited in Kwako, 2003). Considering affective filters, traditional assessment can raise feelings of anxiety, which is powerful enough to considerably limit their performance (Kulm, 1994). As mentioned earlier, in traditional testing situations lead to rote and superficial learning since most of questions are procedural that entail speedy and unreflective responses; accordingly, students are not involved in problem solving and critical thinking (Kwako, 2003).

Initial objectives of the study were to identify Iranian non-English PhD students' perceptions regarding their learning and testing preferences of EPT as an exit exam. Surprisingly, the results obtained from the descriptive analysis revealed that participants perceived deep learning higher than the other subscales. The majority of them believed that EPT introduced students to more useful learning methods in English; allowed them to analyze and produce English; EPT was related to the actual use of the language (writing, speaking, reading, etc.) in English; permitted students to use English in the real world in the future and lastly it improved students' speaking, reading, listening and writing skills. The result is contrary to Kwako's (2003); Thompson and Allen's (2012) and Newstead and Findlay's perspectives (1997) in that high-stake tests encourage students to

concentrate surface and rote approaches to learning rather than deep and meaningful learning as passive learners. Surprisingly, it implies that participants perceive EPT as a measure of understanding instead of memorizing.

Additionally, on the basis of descriptive analyses of learning environment items, 73 participants out of 252 noted that teachers did not play a supportive role in front of students who took classes or preparation courses for this exam; a total of 85 individuals disagreed and strongly disagreed that teachers responded appropriately to their concerns about the test in EPT preparation classes or courses. Moreover, a total of 112 participants agreed assignments designed for preparation classes or courses were clear and 101 of them agreed and strongly agreed assignments designed in preparation classes or courses were useful in this test. In the case of learning outcomes subscale, majority of the participants disagreed and strongly disagreed that EPT could improve their vocabulary and grammar knowledge; interestingly, 142 agreed that the test might improve their translation skills. Descriptive analysis of the surface learning revealed that 140 participants out of 252 confirmed that EPT did not lead to superficial or rote learning. The next aim of the study was to explore whether there was a significant relationship between educational consequences and psychological consequences of EPT among Iranian non-English PhD students or not. The findings presented that there was strong relationship between ECQ and PCQ. In other words, affective factors such as feelings, self-efficacy and motivation were directly pertinent to educational consequences of EPT. Mikami, Leung, and Yoshikawa (2018) considered that psychological factors like motivation, confidence, attitude, and self-esteem are seen as some indispensable parts of learning a second or foreign language and it is even possible that these affective factors are greater in second language testing

circumstances. Shohamy (1982) also affirmed that the affective variables might powerfully impact test takers' success and performance and can lead to erroneous assessment and test anxiety is viewed as a frequently well-known factor that can significantly affect test takers' performance (Yang, 2017).

Furthermore, the effects of high-stake tests on student motivation are undeniable. From Pintrich and Schuffe's studies' responses to open-ended questions motivated when they get involved in an activity since they get pleasure from it or are interested in it; while, students are extrinsically motivated when they engage in an activity as a tool. George (2001) also puts that high-stakes tests are essentially extrinsic motivators for the reason that they make students focus on the outcome or just passing the test. It seems that the most serious result of these negative effects on student motivation is that students may possibly drop out of school (Amrein & Berliner, 2003). It is reported that a number of teachers are even leaving the field of education because of their negative experiences in high-stake testing (Watson, Johanson, Loder, & Dankiw, 2014). Last but not least, self-efficacy has been considered as one of the most powerful variables on students' or learners' achievement and performance in second language learning contexts (Raofi, Tan, & Chan, 2012). Bandura (1997) also confirms that students' beliefs in their abilities influence their performance enormously. Self-efficacy is viewed as a motivational variable in second language learning and it seems almost unfeasible to scrutinize some aspects of individual's functions such as motivation, learning, as well as their academic performance irrespective of the function of learners' self-efficacy beliefs (Pajares & Urdan, 2006).

Taking the limitations of the study into account, the present study was subject to a number of potential methodological weaknesses. For instance, the major limitation of this study was the size of sample under investigation as it was considered as an obstacle in generalizing the results beyond the specific population from which the sample were drawn. Secondly, this study was based on convenience sampling in which participant took part in a study voluntarily; consequently, they probably shared some features not existed in those member of population who were not volunteer and interested to take part in the study. Having access to people, organizations, documents or even sources for whatever reason was considered as a limitation of the current research. Moreover, a number of demographic factors such as gender, age, work experience and socioeconomic status were not controlled and examined in the present study due to restricted access to eligible participants. Another

source of uncertainty that the researcher encountered in the research was the response rate of the questionnaires.

To sum up, the findings of the current research may have noteworthy implications for the understanding of how educational and psychological factors in high-stakes testing can bring effective consequences for the test-takers; thus, the significance of washback awareness among the test developers, language teachers as well as test administrators should not be overlooked and they can utilize their practical knowledge to create new policies regarding high-stakes tests and to think about language testing and assessment in a new way by listening to students voices as marginalized stakeholders in language testing. As a result of this investigation, some suggestions are proposed for future research in this part. For instance, there are still many unanswered questions about the role and effects of some demographic variables including age, gender, and economic status. Increasing the sample size will be required to create better and deep understanding in the next research. There is abundant room for further progress in determining the families, teachers or test developers' voices as other chief stakeholders' to provide a more comprehensive picture of positive and negative consequences of EPT as a high-stakes test in Iran. Last but not least, In future investigations, it might be possible to use a different type of sampling in which participants take part randomly not voluntarily since the possible biases of convenience sampling may be decreased.

References

- Al Amin, M., & Greenwood, J. (2018). The examination system in Bangladesh and its impact: On curriculum, students, teachers and society. *Language Testing in Asia*, 8(4), 1-18. doi: 10.1186/s40468-018-0060-9
- Amrein, A. L., & Berliner, D. C. (2003). The effects of high-stakes testing on students' motivation. *The International Journal of Educational Leadership*, 60(5), 32-38.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman and Company.
- Birenbaum, M., & Feldman, R. (1998). Relationships between learning patterns and attitudes towards two assessment formats. *Educational Research*, 40(1), 90-98. doi:10.1080/0013188980400109
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices* (2nd ed.). White Plains, NY: Pearson Education.

- Brown, J. D. (2008). *Testing in language programs: A comprehensive guide to English language assessment*. Columbus: McGraw-Hill College.
- Castro, M. C., & Vega, S. A. L. (2017). *The impact of tests in students' perceptions and attitudes towards their learning task* (Master's thesis). Universidad Distrital Francisco José de Caldas: Bogotá, Colombia. Retrieved from <http://repository.udistrital.edu.co/bitstream/11349/6612/1/CalderonCastroMelina2017.pdf>
- Dhindsa, H., Omar, K., & Waldrip, B. (2007). Upper secondary bruneian science students' perceptions of assessment. *International Journal of Science Education*, 29(10), 1281-1280. doi: [10.1080/09500690600991149](https://doi.org/10.1080/09500690600991149)
- Dinh, T. P. H. (2020). Washback of English proficiency test in classroom activities at national university of arts education. *VNU Journal of Science: Education Research*, 36 (1), 89-103.
- Estaji, M., & Tajeddin, Z. (2012). The learner factor in washback context: An empirical study investigating the washback of the IELTS academic writing test. *Language Testing in Asia*, 2(1), 5-25.
- Fitzgerald, L. (2015). *Consequences of high-stake testing* (Master's thesis). Rochester, NY: St. John Fisher College, Fisher Digital Publications. Retrieved from http://fisherpub.sjfc.edu/education_ETD_masters
- Genesee, F., & Hamayan, E. V. (1994). Classroom-based assessment. In F. Genesee (Eds.), *Educating second language children* (pp.212-239). New York: Cambridge University Press.
- George, P. S. (2001). A + accountability in Florida? *Educational Leadership*, 59(1), 28-32.
- Gilmore, A., & Smith, J. (2008). *Students' experience of assessment*. Retrieved from https://www.researchgate.net/publication/267301534_Students%27_Experience_of_Assessment.
- Gipps, C. V. (2012). *Beyond testing: Towards a theory of educational assessment*. London: Routledge.
- Gunn, J., Al-Bataineh, A., & Abu Al-Rub, M. (2016). Teachers' perceptions of high-stakes testing. *International Journal of Teaching and Education*, 6(2), 49-62. doi: 10.20472/TE.2016.4.2.003
- Hung, S. A., & Huang, H. D. (2019). Standardized proficiency tests in a campus-wide English curriculum: A washback study. *Language Testing in Asia*, 9 (21). doi:[10.1186/s40468-019-0096-5](https://doi.org/10.1186/s40468-019-0096-5)
- Kheirkhah, H., & Ghonsooly, B. (2014). Qualitative study of Iranian English university entrance examination in the light of positive washback strategies. *Studies in English Language Teaching*, 2(1), 37-65. doi:[10.22158/selt.v2n1p38](https://doi.org/10.22158/selt.v2n1p38)
- Khodabakhshzadeh, H., Zardkanloo, R., & Alipoor, I. (2017). The effects of mock tests on Iranian EFL learners' test scores. *International Journal of Education & Literacy Studies*, 5(3), 47-51.
- Khoshsima, H., Saed, A., & Mousaei, F. (2018). Exploring the effect of teaching test-taking strategies on intermediate level learners on reading section of IELTS: Learners' attitude in focus. *Advances in Language and Literary Studies*, 9(2), 4-9.
- Kulm, G. (1994). *Mathematics assessment: What works in the classroom?* San Francisco, CA: Jossey Bass Inc.
- Kwako, J. (2003). *A brief summary of traditional and alternative assessment in the college classroom*. Retrieved from www.stat.wisc.edu/~nordheim/Kwako_assessment4.doc
- Law, B., & Eckes, M. (2007). *Assessment and ESL: An alternative approach* (2nd ed.). Canada: Portage & Main Press. Retrieved from <https://books.google.com/books>
- Menken, K. (2017). High-stakes tests as de facto language education policies. In E. Shohamy, I.G. Or & S. May (Eds.), *Language testing and assessment. Encyclopedia of language and education* (pp. 385 -397). Cham, Switzerland: Springer. doi: 10.1007/978-3-319-02261-1_35
- Mikami, H., Leung, C.Y., & Yoshikawa, L. (2018). The threshold of anxiety in low-stakes testing for foreign language reading. *Reading in a Foreign Language*, 30(1), 92-107.
- Munoz, R. E. (2017). The effect of washback on EFL learners' attitude toward tests. *Studies in English Language Teaching*, 5(3), 516-530.
- Newstead, S. E., & Findlay, K. (1997). Some problems with using examination performance as a measure of teaching ability. *Psychology Teaching Review*, 6(1), 23-30.
- Ostovar Namaghi, S. A. (2013). Washback from bottom-up: A grounded theory. *International Journal of Applied Linguistics and English Literature*, 2(6), 213-220.
- Pajares, F., & Urdan, T. (2006). Foreword. In F. Pajares, & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 9-12). Greenwich, CT: Information Age Publishing.
- Pan, Y. C. (2009b). Test impact: English certification exit requirements in Taiwan. *TEFLIN Journal*, 20(2), 119-139.
- Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in education: Theory, research, and applications*. Upper Saddle River, NJ: Merrill Prentice Hall.
- Poehner, M. E. (2013). Dynamic assessment in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp.3-4). Oxford: Blackwell Publishing Ltd. doi: 10.1002/9781405198431.wbeal0345
- Raoofi, S., Tan, B. H., & Chan, S. H. (2012). Self-efficacy in second/foreign language learning contexts. *English Language Teaching*, 5(11), 60-74. doi:[10.5539/elt.v5n11p60](https://doi.org/10.5539/elt.v5n11p60)
- Rezaeian, M., Seyyedrezaei, S. H., Barani, G., & Seyyedrezaei, Z. (2020). Construction and validation of educational, social and psychological consequences questionnaires of EPT as a high-stakes test. *International Journal of Language Testing*, 10(2), 33-70.
- Rezaeian, M., Seyyedrezaei, S. H., & Seyyedrezaei, Z. (in press). A study of motivation, self-efficacy beliefs and feelings as psychological factors among Iranian non-English PhD students in EPT. *The Iranian Journal of Learning and Memory*. doi: [10.22034/iepa.2020.243321.1197](https://doi.org/10.22034/iepa.2020.243321.1197)

- Salehi, M. [Massomeh], & Marefat, F. (2014). The effects of foreign language anxiety and test anxiety on foreign language test performance. *Theory and Practice in Language Studies*, 4(5), 931-940. doi:10.4304/tpls.4.5.931-940
- Salehi, M. [Mohammad], & Tarjoman, M. (2017). An investigation of a nationwide exam from a critical language testing perspective. *Cogent Social Sciences*, 3(1). doi:10.1080/23311886.2017.1396639
- Shohamy, E. (1982). Affective considerations in language testing. *The Modern Language Journal*, 66(1), 13-17. doi:10.2307/327810
- Shohamy, E. (2001b). Democratic assessment as an alternative. *Language Testing*, 18(4), 373-391. doi:10.1177/026553220101800404
- Shohamy, E. (2007b). Tests as power tools: Looking back, looking forward. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. E. Turner & C. H. Doe (Eds.), *Language testing reconsidered* (pp.141-153). Ottawa: University of Ottawa Press.
- Shohamy, E. (2017). Critical language testing. In E. Shohamy, I. G. Or & S. May (Eds.), *Language testing and assessment, encyclopedia of language and education* (pp.441-455). Springer International Publishing AG. doi: 10.1007/978-3-319-02261-1_35
- Siahpoosh, H., Ramak, M., & Javandel, B. (2019). Washback effect of IELTS on Iranian learners' perspectives on IELTS preparation courses. *Journal of Applied Linguistics and Language Research*, 6(2), 43-52.
- Struyven, K., Dochy, F., & Oblermann, H. (2012). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment & Evaluation in Higher Education*, 30(4), 331-347. doi: 10.1080/0260293042000318091
- Thibodeaux, A. K. (2014). *The effects of leadership and high-stakes testing on the retention of teachers* (Doctorial Dissertation). University of Southern Mississippi, Hattiesburg.
- Thompson, G., & Allen, T. (2012). Four effects of the high-stakes testing movement on African American K-12 students. *The Journal of Negro Education*, 81(3), 218-227.
- Watson, C., Johanson, M., Loder, M., & Dankiw, J. (2014). Effects of high-stakes testing on third through fifth grade students: Student voices and concerns for educational leaders. *Journal of Organizational Learning and Leadership*, 12(1), 1-11.
- Wheelock, A., Bebell, D. J., & Haney, W. (2000). *What can student drawings tell us about high-stakes testing in Massachusetts? Teachers College Record*. Retrieved from <https://www.tcrecord.org/library>
- Yang, Y. X. (2017). Test anxiety analysis of Chinese college students in computer-based spoken English test. *Educational Technology & Society*, 20(2), 63-73.

Appendix I: Educational Consequences Questionnaire (English Version)

Dear Respondent,

We would like to ask you to help us by answering the following questions concerning EPT as a high-stakes test. This survey is conducted with the aim of looking into PhD non-English students' perceptions of the educational consequences of EPT. This is not a test so there is no "right" or "wrong" answers and you don't even have to write your name on it. We are interested in your personal opinion. The information will be kept confidential and will be used just for research purposes. Please give your answers sincerely as only this will guarantee the success of the investigation. Thank you very much for your help and cooperation.

1. Strongly agree 2. Agree 3. Undecided 4. Disagree 5. Strongly disagree

Constructs	Strongly agree	Agree	Undecided	Disagree	Strongly disagree
Learning Environment					
Instructors play a supportive role for students participating in test preparation classes / courses.					
Instructors provide appropriate answers to students' concerns about this test in preparation classes / courses.					
Assignments designed for EPT preparation classes or courses are clear					
Assignments designed for EPT preparation classes or courses are useful.					
Surface learning	Strongly agree	Agree	Undecided	Disagree	Strongly disagree
5. This test makes students memorize test-related contents.					
6. This test increases superficial learning among students.					
Deep Learning	Strongly agree	Agree	Undecided	Disagree	Strongly disagree
7. This test makes students learn more useful ways to learn English.					
8. This test makes students analyze English.					
9. This test makes students produce English.					
10. This test helps students to better understand English.					
11. The questions of this test are related to the actual use of the language (writing, speaking, reading, etc.) English.					
12. This test makes students use English in a real environment in the future.					
13. This test improves students' speaking skills.					
14. This test improves students' reading skills.					
15. This test improves students' listening skills.					
16. This test improves students' writing skills.					
Learning Outcomes	Strongly agree	Agree	Undecided	Disagree	Strongly disagree
17. This test improves students' vocabulary.					
18. This test improves students' grammar.					
19. This test improves students' translation skills.					
Testing Preferences	Strongly agree	Agree	Undecided	Disagree	Strongly disagree
20. I prefer all language skills, including speaking and writing, to be assessed in this test.					
21. I prefer questions that require more or deeper thinking.					
22. I prefer questions that require creativity to be answered.					
23. I prefer questions that need further analysis.					

Appendix II: Psychological Consequences Questionnaire (English Version)

Dear Respondent,

We would like to ask you to help us by answering the following questions concerning EPT as a high-stakes test. This survey is conducted with the aim of looking into PhD non-English students' perceptions of the psychological consequences of EPT. This is not a test so there is no "right" or "wrong" answers and you don't even have to write your name on it. We are interested in your personal opinion. The information will be kept confidential and will be used just for research purposes. Please give your answers sincerely as only this will guarantee the success of the investigation. Thank you very much for your help and cooperation.

1. Strongly agree 2. Agree 3. Undecided 4. Disagree 5. Strongly disagree

Constructs	Strongly agree	Agree	Undecided	Disagree	Strongly disagree
Self-Efficacy Beliefs					
1. I believe I have the ability to pass this test.					
2. I'm sure I can improve my English with more effort.					
3. I think learning English is very easy.					
Feelings					
4. This test increases my stress and anxiety.					
5. The result of this test makes me feel hopeless.					
6. The result of this test makes me nervous / aggressive.					
7. This test has caused stress and tension in my family.					
8. The difficulty of preparing for this test will dampen my motivation to study English in the future.					
9. This test makes students think about dropping out of university.					
Intrinsic Motivation					
10. Preparing for the test increases my interest in learning English.					
11. Preparing for this test makes the learning process enjoyable for me.					
12. Preparing for this test makes me aware of the value of learning English.					
13. Preparing for this test gives me inner satisfaction.					
Extrinsic Motivation					
14. I want to pass this exam to get a job promotion.					
15. Making a distinction among the participants based on the result of this test will increase my motivation to study.					

How to Site: Rezaeian, M., Seyyedrezaei, S., Barani, G., & Seyyedrezaei, Z. (2020). An Investigation into Iranian non-English DDDddd'''' '''' eennnnn mnnnnnn eiiii gg ss nn etttt tt ll oosseueeee ff PP.. *Iranian Journal of Learning and Memory*, 3(10), 69-79. doi: 10.22034/iepa.2020.235057.1176



Iranian Journal of Learning & Memory is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.