The Washback Effect of Task-based Assessment on the Iranian EFL Learners' Development of Pragmatic Competence

Maryam Nazari, Ph.D. Candidate, Department of English, Malayer Branch, Islamic Azad University, Malayer, Iran
yasinmydiamond6@gmail.com
Abbas Bayati*, Assistant Professor, Department of English, Malayer Branch, Islamic Azad University, Malayer, Iran
abbasbayat305@gmail.com
Payman Rajabi**,** Assistant Professor, Department of English, Malayer Branch, Islamic Azad University, Malayer, Iran
paymanrajabi2002@gmail.com

Abstract
The present study was an attempt to explore the 'washback effect' of task-based assessment (TBLA) on EFL Iranian learners' pragmatic development. To this end, through conducting KET (Key English Test), 60 out of 120 EFL Iranian learners studying in an English language school, were randomly selected. They were assigned to treatment group (N=30), and control group (N=30). The treatment group was assessed through TBLA and the control group was assessed via non-task based assessment for 20 ninety-minute sessions. The class sessions were held twice a week. The obtained data was examined through independent sample *t*-test. The findings implied that TBLA as a pedagogical measurement tool can well replace the classic assessment procedures, since all educational efforts including testing and assessment procedures are planned to maximize the educational gains and developments.

Keywords*:* Task-based language assessment, washback effect, pragmatic competence, speech acts

Introduction

Testing is one of the main concerns of the trained teachers. It is a complex process and very important in education because as Chastain (1988) states, it "measures more material over longer periods of time, thereby stressing a greater degree of retention, organization, and comprehension of the material" (p. 379). Thus, it is evident that testing has influence on the instruction process. This influence of testing on teaching and learning is called 'the washback effect' in language testing (Alderson & Wall, 1993; Gates, 1995; Hughes, 1989; Pan, 2009).
Wesche (1983) points says:

By making our tests more reflective of the kinds of situations, language content, and purposes for which second language speakers will need their skills, we will be able to make more accurate predictions about how they will be able to function using the target language in real life. She further notes that such testing is likely to have dramatic effects on the format and content of second language curricula as well and to improve student motivation through its increased relevance". (p. 53)

According to Boyle, J. & Falvey, P. (1994), 'introducing tests at every point into an educational system leads to many intentional or unintentional changes in the curriculum, especially in teaching and learning practices. They consider washback as one of the Big Four considerations in evaluating the worth of a test. In fact, tests are sometimes perceived as exerting

a power which hinders progress in learning, and sometimes as Fullilove, (1994) believes ‘not only have many tests failed to change, but they have continued to exert a powerful negative washbach effect on teaching" (p. 57). They also note that ‘educationalists often decry the negative washback effects of examinations and regard washback as an impediment to educational reform or progressive education’ (p. 60).

Cronbach (1971) believed that validity is not a property of the test or assessment but rather of the meaning of the test scores. Therefore, what is to be validated is not the test or observation per se but rather the influences derived from test scores or other indicators inferences about score meaning or interpretation entails.

Literature Review

According to Messick, (1996), authentic assessments pose engaging tasks in realistic settings or close simulations so that the tasks and processes, as well as available time and resources, parallel those in the real world. He defines direct assessments as open-ended tasks in which the respondent can freely perform the complex skill at issue not restricted by structured item forms restrictive response formats. The goal, he adds, is to minimize constraints on examinee behavior associated with sources of construct-irrelevant method variance such as test-wiseness in coping with various item types, differential tendencies toward guessing, and other artificial restrictions on examinees' representations of problems and on their modes of thinking or response. Assessment, according to Şenel and Tütüniş (2013), is ‘an ongoing process that encompasses a much wider domain. Whenever a student responds to a question, offers a comment, or tries out a new word or structure, the teacher subconsciously makes an assessment of the student’s performance’.

To achieve the above mentioned goal, performance-based assessments may be recommended. Performance-based assessments "represent a set of strategies for the application of knowledge, skills, and work habits through the performance of tasks that are meaningful and engaging to students" (Hibbard, 1996, p. 5). Because authentic tasks are rooted in curriculum, tasks can be developed based on what already works for a special context. Through this process, assignments become more authentic and more meaningful to students. Tavakoli and Skehan (2005) presented a model of task-based performance in relation to language testing, the main purpose of which is to make obvious that the rating assigned to someone on the basis of their performance on a task is the consequence of a whole range of factors, only one of which can possibly be their underlying competence.

The difference between communicative tests and task-based tests is that a communicative speaking test, for example, might involve comparing two pictures or talking about one’s opinions on vegetarianism, while a task-based test would only use such formats in the unlikely situation where detailed visual descriptions or defending one’s views on vegetarianism were communicative contexts test takers would encounter in their professional or academic life outside of the classroom (Fischer, et al., 2011). It is, thus, evident that language testers select a special type of test to assess learners based on the goal that is going to be achieved.

As for pragmatics, it is a subfield of linguistics that has been defined as “the study of language from the point of view of users, especially of the choices they make, the constraints they encounter in using language in social interaction and the effects their use of language has on other participants in the act of communication” (Crystal, 1997, p. 301). As cited in Motallebi Nia, Z. (2013), Bachman in his 1990 model, is the first one to include pragmatic competence as a separate and significant component of communicative competence. Pragmatic competence, he states, refers to "the knowledge of communicative action and how to carry it out" and "the ability

to use the language appropriately according to the context" (p. 87). Currently, this term is extensively used in the field of SL and FL acquisition and teaching, especially in reference to pragmatic competence as one of the abilities subsumed by the overarching concept of communicative competence. The notion of *pragmatic competence* was first defined by Chomsky (1980) as the "knowledge of conditions and manner of appropriate use (of the language), in conformity with various purposes" (p.224). This concept was seen in opposition to *grammatical competence* that in Chomskyan terms is "the knowledge of form and meaning". In a more contextualized fashion, Canale & Swain (1980) included *pragmatic competence* as one important component of their model of *communicative competence*. In this model, *pragmatic competence* was identified as *sociolinguistic competence* and defined as the knowledge of contextually appropriate language use (Canale & Swain, 1980; Canale, 1988).

Pragmatic competence may be considered as a fundamental aspect of a more general communicative competence. In language teaching, communicative competence is defined as the students' ability to "understand the essential points of what a native speaker says… in a real communicative situation" as well as "respond in such a way that the native speaker interpret to response with little or no effort and without errors that are so distracting that they interfere drastically with communication" (Terrell, 1977, p. 326, cited in Rajabi, Azizifar & Gowhary, 2015).

Based on the above background, this study aimed to compare TBLA and non-task assessment in terms of their washback effect on the development of pragmatic competence of Iranian EFL learners. In fact, despite numerous studies regarding task-based teaching and learning as well as assessment through task, research is still lacking on the washback effect of task-based assessment on Iranian EFL learners' pragmatic development. Thus, this study was conducted to investigate if there is any relation between task-based assessment and pragmatic development. To achieve the purposes of the study, the following research questions and hypotheses were formulated:

Q1. Does the washback effect of task-based assessment affect EFL learners' pragmatic competence?

Q2. Is there any significant difference between the non-task language assessment and TBLA concerning their washback effect on EFL learners' pragmatic development?

Q3. How does washback affect the pragmatic development achieved through task-based or non-task assessment?

Ho1. The washback effect of task-based assessment does not influence pragmatic competence.

Ho2. There is no significant difference between the non-task language assessment and task-based assessment concerning their washback effect on EFL learners' pragmatic development.

Ho3. Washback does not influence the pragmatic development achieved through task-based or non-task assessment.

Method

Design

This study was conducted through a pretest, posttest equivalent-group design, with 2 groups (treatment and control) of equal size acting as participants.

Participants

The participants of the study were 60 EFL learners of Kermanshah Jahad Daneshgahi Language School, selected from a pool of 120 students. For the sake homogeneity in terms of language proficiency, KET (Key English Test) was administered to the pool and based on the KET scores, 60 participants who scored +1 SD from the mean were chosen as pre-intermediate level EFL learning sample of the study. They were randomly divided into two groups (treatment and control) of the same size. Their age range was 13 to 35, and all of them were school or university students speaking Persian as their first language.

Material

The material used in this study consisted of the speech acts extracted from *Top Notch* book series.

Instruments

The testing instruments used in this study for different purposes were the sample KET (Key English Test), six teacher-made task-based pragmatic quizzes, six non-task multiple-choice pragmatic quizzes, and a 42-item pragmatic test as pre- and posttests.

Ket tests the four skills of reading, writing, listening and speaking and is based on the Waystage specification (1990, Council of Europe). The six task-based pragmatic quizzes (for the treatment group) and six non-task pragmatic quizzes (for control group) were administered every 2 sessions. There were nearly 10 multiple-choice test items in each quiz for control group and some pragmatic task items for the treatment group. The 42-item pragmatic test (pretest) consisted of 42 multiple-choice test items and was testing the participants' knowledge of pragmatics and speech acts. The same 42-item multiple choice pragmatic test (pretest) was administered to both groups as the posttest after the treatment. Moreover, the researcher considered a period of instruction for the teaching of the intended pragmatic points to the two groups.

Procedures

To come up with the most appropriate answer to the research questions, collecting, describing and analyzing the data was indispensable. Thus, the following steps were taken. In the first stage 120 EFL learners placed at the same level courses were selected from among the EFL learners of Jahad Daneshgahi language school in Kermanshah. Then, a KET sample test (2010) was administered to determine the participants' true level of language proficiency. Next, 60 students who scored within +1 SD from the mean were chosen as the participating sample of the study and were randomly put into the treatment and control groups, each with 30 participants.

In the second stage, both treatment and control groups took the pragmatic pretest as a measure of the participants' pragmatic knowledge of the selected English speech acts. In the third stage, both groups were independently taught the speech acts in the same way using *Top Notch* book series, for 20 sessions of ninety minutes. The class sessions were held twice a week.

To nullify the potential effect of methodology, both groups were taught by the same instructor. After every two sessions, the researchers administered a quiz to both groups. The treatment group was given the task-based quiz and the control group was assessed through classic method (multiple-choice here). In this manner, the researchers administered six separate quizzes during twenty sessions and tried to find out the possible differential effect of the two assessment types on the pragmatic development and final achievement of the participants in both groups. Finally, the posttest was administered to the two groups after the treatment sessions were over.

Results and Discussion

Before the main study, as was mentioned above, the participants in the two groups were compared in terms of homogeneity of their general language ability. To do so, an independent samples *t*-test was used. The table below shows the related descriptive statistics:

Table 1. *Descriptive Statistics for the treatment and control groups*

	Groups	N	Mean	Std. Deviation
KET	Treatment group	30	83.47	26.57
	Control group	30	80.00	29.67

The results of the *t*-test show that the difference between the two groups in terms of their general English language ability was not statistically significant. Thus, the two groups (the treatment and control groups) were homogeneous in terms of English language proficiency. It is to be noted that the participants with extreme scores were excluded from the study, although they attended the classes.

Table 2. *Independent Samples Test results for KET scores*

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
KET	Equal assumed	v.413	.523	.477	58	.635	3.46	7.27	-11.09	18.02
	Equal not assumed	v		.477	57.3	.635	3.46	7.27	-11.09	18.02

As shown in the above table, the difference between the groups was not statistically significant; thus, the participants in the two groups were homogeneous in their general language ability. Accordingly, the researcher could safely start the treatment.

The homogeneity of the participants in terms of their pragmatic knowledge was also checked, that is, a pragmatic pretest was run (an independent samples *t*-test) to detect whether the statistical difference between the means of the two groups in this construct was significant or not. The pragmatic test consisted of 40 multiple-choice which measured the participants' knowledge of pragmatics and speech acts. Table 3 below shows the descriptive statistics:

Table 3. *Descriptive Statistics for the two groups in the pretest*

	Groups	N	Mean	Std. Deviation	Std. Error Mean
Pretest	Treatment group	30	12.00	3.57	.65
	Control group	30	13.27	4.01	.73

Table 4. *Independent Samples t-test results for the pragmatic pretest*

		Levene's Test Equality Variances		fort-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Pretest	Equal v. assumed	.522	.473	-1.291	58	.202	-1.267	.981	-3.23	.69
	Equal v not assumed			-1.291	57.2	.202	-1.267	.981	-3.23	.69

As tables 3 and 4 show, the difference between the two groups regarding their performance in the pragmatic pretest was not statistically significant. Thus, the treatment could safely be done. The treatment was of two types: presenting task-based tests to the treatment group, and presenting non-task tests to the control group, while using common traditional techniques of teaching.

As for posttest results, after the treatment finished in 16 sessions, the posttest mean scores of the two groups were compared to see if they had shown a significant difference in their pragmatic knowledge, thereby finding answer to the first research question. To this purpose, the mean scores of the six tests in the treatment group were compared in pairs, with the following results:

Table 5. *Paired Samples Statistics for the Progress Tests in Treatment Group*

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Exptest1	13.47	30	4.133	.755
	Exptest2	14.87	30	2.897	.529
Pair 2	Exptest2	14.87	30	2.897	.529
	Exptest3	14.90	30	3.133	.572
Pair 3	Exptest3	14.90	30	3.133	.572
	Exptest4	14.77	30	4.384	.800
Pair 4	Exptest4	14.77	30	4.384	.800
	Exptest5	16.63	30	2.965	.541
Pair 5	Exptest5	16.63	30	2.965	.541
	Exptest6	16.77	30	2.897	.529

In table 5, the comparison of the mean scores in the progress tests respectively shows that language learners improved their pragmatic competence in these tests. The gradual increase in their mean scores indicates that the administration of these tests one after the other caused a mellow increase in their mean scores.

Table 6. *Paired Samples Correlations between pairs of tests*

		N	Correlation	Sig.
Pair 1	Exptest1 & Exptest2	30	.662	.000
Pair 2	Exptest2 & Exptest3	30	.754	.000
Pair 3	Exptest3 & Exptest4	30	.445	.014
Pair 4	Exptest4 & Exptest5	30	.516	.004
Pair 5	Exptest5 & Exptest6	30	.893	.000

The correlation between each pair of tests is shown in table 6 above. The mean scores of each pair of tests are compared with one another respectively. As shown in table 7 below, the difference of mean scores in the different intervals is not statistically significant. Although there was an increase in the mean scores of these progress tests during the study, statistically such difference in most cases was not significant.

Table 7. *Paired Samples Test Results for the Progress Tests in the Treatment Group*

		Paired Differences					T	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Exptest1 Exptest2	-1.400	3.103	.566	-2.559	-.241	-2.471	29	.020
Pair 2	Exptest2 Exptest3	-.033	2.125	.388	-.827	.760	-.086	29	.932
Pair 3	Exptest3 Exptest4	.133	4.100	.749	-1.398	1.664	.178	29	.860
Pair 4	Exptest4 Exptest5	-1.867	3.821	.698	-3.294	-.440	-2.676	29	.012
Pair 5	Exptest5 Exptest6	-.133	1.358	.248	-.640	.374	-.538	29	.595

The same comparison was done between the mean scores of the control group in terms of the paired tests administered to them. The results are shown in table 8 below:

Table 8. *Paired Samples Statistics for pairs of tests in the Control Group*

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Controltest1	13.17	30	3.975	.726
	Controltest2	13.17	30	3.152	.576
Pair 2	Controltest2	13.17	30	3.152	.576
	Controltest3	13.80	30	3.614	.660
Pair 3	Controltest3	13.80	30	3.614	.660
	Controltest4	14.63	30	3.419	.624
Pair 4	Controltest4	14.63	30	3.419	.624
	Controltest5	13.97	30	3.499	.639
Pair 5	Controltest5	13.97	30	3.499	.639
	Controltest6	14.10	30	3.556	.649

The comparison shows that gradual increase in the learners' mean scores did not exist in the control group. Actually, the learners showed more or less the same mean score in their six progress tests. This mean score was somehow between 13 and 14.50.

Table 9. *Paired Samples Correlations between pairs of tests in the Control Group*

		N	Correlation	Sig.
Pair 1	Controltest1 & Controltest2	30	.867	.000
Pair 2	Controltest2 & Controltest3	30	.617	.000
Pair 3	Controltest3 & Controltest4	30	.588	.001
Pair 4	Controltest4 & Controltest5	30	.699	.000
Pair 5	Controltest5 & Controltest6	30	.760	.000

Here again, the mean scores of each pair of tests are compared with one another respectively. As shown in Table 10 below, the difference of mean scores in the different intervals is not statistically significant. As there was no increase in the mean scores of these progress tests during the study, such difference in most cases was not statistically significant. Accordingly, the first null hypothesis that, 'The washback effect of task-based assessment does not influence pragmatic competence' is not rejected, meaning that task-based assessment did not have washback effect on the learners' performance in the posttest.

Table 10. *Paired Samples Test Results for Pairs of Tests in the Control Group*

	Paired Differences	t	df	Sig. (2-

		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			tailed)
Pair 1	Controltest1 Controltest2	-.000	2.000	.365	-.747	.747	.000	29	1.000
Pair 2	Controltest2 Controltest3	--.633	2.988	.546	-1.749	.483	-1.161	29	.255
Pair 3	Controltest3 Controltest4	--.833	3.196	.583	-2.027	.360	-1.428	29	.164
Pair 4	Controltest4 Controltest5	-.667	2.682	.490	-.335	1.668	1.361	29	.184
Pair 5	Controltest5 Controltest6	--.133	2.446	.447	-1.047	.780	-.299	29	.767

To check if using non-task language assessment and TBLT show a significant difference concerning their washback effect on EFL learners' pragmatic development, an independent samples *t*-test was used and the mean scores of the treatment group and the control group were compared. The results appear in the following table:

Table 11. *Descriptive Statistics for Posttest Scores in Treatment and Control Groups*

	Groups	N	Mean	Std. Deviation	Std. Error Mean
Posttest	Treatment group	30	17.37	2.59	.47
	Control group	30	14.63	3.37	.61

As shown in table 11 above, the mean and standard deviation of the treatment group (M=17.37, SD=2.59), are different from those of the control group (M=14.63, SD=3.37). Thus, to compare the mean scores of these two groups, an independent samples *t*-test was run.

In order to detect the difference between the two groups in terms of their performance in the posttest, an independent samples t-test was run. The results in table 12 below show that there was a statistically significant difference between the two groups with regard to their posttest scores, t (58) = 3.515, p = 0.001.

Table 12. *Independent Samples Test Results for Comparing the Posttest Scores in Treatment and Control Groups*

	Levene's for Equality of Variances	Testt-test for Equality of Means

		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
Posttest	Equal assumed	v4.095	.048	3.51558		.001	2.73	.77	1.17	4.29
	Equal v not assumed			3.51554.3		.001	2.73	.77	1.17	4.29

Accordingly, the second research null hypothesis, that is, 'There is no significant difference between the non-task language assessment and task-based assessment concerning their washback effect on EFL learners' pragmatic development' was rejected, meaning that those participants who benefited from task-based assessment during the study outperformed the control group in their performance in the pragmatic knowledge posttest.

As for the third research question, to find a reasonable answer, a paired samples *t*-test was run to compare the posttest mean scores of the treatment and control groups. Table 13 below shows the related descriptive statistics:

Table 13. *Paired Samples Statistics for Treatment and Control Groups in pretest and posttest*

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	ExpGpretest	12.00	30	3.57	.65
	ExpGposttest	17.37	30	2.59	.47
Pair 2	ContGpretest	13.27	30	4.01	.73
	ContGposttest	14.63	30	3.37	.61

The intra-group comparison of the mean scores shows that the gap between the mean scores in the treatment group is more than that of the control group.

In order to check whether there was a statistically significant difference between the pretst and posttest in each group, a paired-samples *t*-test was run. As shown in Table 14 below, there was a significant difference between the pretest and posttest in each group with regard to their pragmatic competence. Since both task-based assessment and non-task assessment made a significant difference in the learners' performance, it can be inferred that the washback of these two types of assessment was the same, and thus task-based assessment does not influence the pragmatic development more differently than non-task assessment.

Table 14. *Paired Samples Test Results for Comparing the Posttest Scores in Treatment & control Group*

	Paired Differences				t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference			

				Lower	Upper			
Pair 1	ExpGpretest ExpGposttest	--5.36	3.66 .66	-6.73	-3.99	-8.02	29	.000
Pair 2	ContGpretest ContGposttest	--1.36	2.74 .50	-2.39	-.34	-2.72	29	.011

Based on the above discussion, the third research null hypothesis, that is, 'Washback does not influence the pragmatic development achieved through task-based or non-task assessment' is rejected, meaning that washback of both task-based and non-task assessments significantly influence the learners' pragmatic competence.

Conclusions

This study aimed at investigating the washback effect of task-based assessment on the Iranian EFL learners' development of pragmatic competence. To this purpose, based on the data elicited from the EFL learners in Kermanshah Jahad Daneshgahi Language School, the three formulated research null hypotheses were tested. In fact, since the study was exposed factor in nature, the effect of the independent variable, that is, the type of assessment, on the dependent variable; meaning, learners' pragmatic competence was explored. Thus, in the treatment group, the task-based assessment in the form of presenting six progress task-based tests was utilized, while in the control group, non-task tests were used. Having implemented the assessment for 6 sessions during the study, the learners' performance was measured through the posttest, and statistical data analysis was done. As the results of descriptive statistics revealed, the first null hypothesis related to the washback effect of task-based assessment was not rejected, while the second and third null hypotheses mentioned above were rejected. The findings of this research, derived from the analysis of results, may contribute to a better understanding of EFL teachers concerning the issues related to washback impact of task-based assessment, and may assist them in enhancing learners' pragmatic competence.

References

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14, 115-129.

Backman, L. F., & Palmer, A. S. (1996). *Language testing in practice.* Oxford: Oxford University Press.

Bachman, LF, & Palmer, AS (2010). *Language assessment in practice: developing language tests and justifying their use in the real world.* Oxford: Oxford University Press.

Bagherkazemi, M. (2014). *Interlanguage pragmatic Development: Impacts of Individual Output, Collaborative Output, Input Enhancement, Metapragmatic Awareness Raising, and Pragmatic Learning Strategies.* Ph.D dissertation. Allameh Tabatabaei University. Tehran, Iran.

Boyle, J. & Falvey, P. (1994). *English language testing in Hong Kong.* Hong Kong Chinese University Press.

Canale, M. (1983). From communicative competence to language pedagogy, in J. Richards and R. Schmidt (eds.).

Canale, M. (1988). The measurement of communicative competence. *Annual review of applied linguistics*, 8, 67-84.

Canale, M. and M. Swain. (1980). Theoretical bases of Communicative approaches to second language testing. *Applied Linguistics* 1: 1-47.

Chastain, K. (*1988*). *Developing second language skills Theory and practice (3rd ed.).* Florida Harcourt Brace Jovanovich.

Chomsky, N. (1980). *Rules and representations.* New York: Columbia University Press.

Crystal, D. (1997). *A dictionary of linguistics and phonetics. 2nd edition.* Oxford: Blackwell.

Ellis, R. (2003). *Task-based Language Learning and Teaching*. Oxford University Press. ISBN 0-19-442159-7

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning*, 54(2), 227-275.

Ellis, R. (2008). *The Study of Second Language Acquisition (2nd Ed.).* Oxford Oxford University Press.

Fischer, J. & et. al. (2011). *Guidelines for task-based university language testing* http://book.coe.int Council of Europe Publishing

Fullilove, J. (1994). *Assessing spoken English in public examinations — why and how?* J. Boyle, P. Falvey (Eds.)

Gates, S. (1995). *Exploiting washback from standardized tests.* In J. D. Brown & S. O. Yamashita (Eds.), *Language testing in Japan* (pp. 101–106). Tokyo: Japanese Asso-ciation for Language Teaching.

Hibbard, K. (1996). *Performance-Based Learning and Assessment. A Teacher's Guide.*

Hughes, A. (1989). *Testing for language teachers*. Cambridge and New York: Cambridge University Press ix + 172 pp. ISBN 0-521-27260-2

Kasper, G. (1997). *Can pragmatic competence be taught?* NetWork, 6, 105e119.

Kasper, G., & Rose, K. (2002). *Pragmatic development in a second language.* Oxford, UK: Blackwell.

Leech, G (2014). *The pragmatics of politeness.* Oxford: Oxford University Press.

McNamara, T. (2001). Language assessment as social practice: Challenges for research. *Language Testing*, 18(4), 333–350.

Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13, 241–256.

Motallebi Nia (2013). *The Washback of English Tests to EFL Teacher and Learner Beliefs and Practices Concerning Pragmatics.* M.A. thesis. Allameh Tabataba'I University. Tehran, Iran.

Pan, Y.-C. (2009). A review of washback and its pedagogical implications. *VNU Journal of Science, Foreign Languages*, 25, 257–263.

Purpura, J (2004). *Assessing grammar*. Cambridge: Cambridge University Press.

Rajabi, Sh., Azizifar, A., & Gowhary, H. (2015). The effect of explicit instruction on pragmatic competence development; teaching requests to EFL learners of English. *Social and Behavioral sciences* 199: 231-239.

Roever, C, Fraser, C, Elder, C (2014). *Testing ESL sociopragmatics: development and validation of a web-based test battery.* Frankfurt: Peter Lang.

Şenel, E. & Tütüniş, B. (2013). *The Washback Effect of Testing on Students' Learning In EFL Writing Classes.* https://www.researchgate.net/publication/237061539.

Taguchi, N, & Roever, C (2017). *Second language pragmatics.* Oxford: Oxford University Press.

Tavakoli, H. (2012). *A dictionary of research methodology and statistics in applied linguistics*. Rahnama press.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In Ellis (Ed.), *planning and task performance in a second language*. Amsterdam: John Benjamines.

Terrell, T. D. (1977). *A Natural Approach to Second Language Acquisition and Learning.*

Timpe, V (2013). *Assessing intercultural language learning.* Frankfurt: Peter Lang.

Timpe-Laughlin, V., Wain, J., & Schmidgall, J. (2015). *Defining and operationalizing the construct of pragmatic competence: review and recommendations.* ETS Research Report Series 2015 (1), 1–43. Princeton, NJ: ETS.

Wesche, M. B. (1983). *Communicative Testing in a Second Language.* https://doi.org/10.1111/j.1540-4781.1983.tb02499.x.

Youn, S. (2015). *Validity argument for assessing L2 pragmatics in interaction using mixed methods.* Language Testing, 32(3), 199–225.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی