

فصلنامه علمی - پژوهشی علوم انسانی دانشگاه الزهراء (س)

سال هجدهم، شماره ۷۰، بهار ۱۳۸۷

The Washback Effect Of Alternatives In Assessment On Iranian EFL Pre-University Learners' Achievement

F. Heidari¹

Abstract

This study investigation the washback effect of alternatives in assessment (i. e. self-assessment, cross-level peer-assessment, same-level peer-assessment, and conference) on Iranian EFL learners. To achieve the purpose of this study, a pre-test, Post-test, and four formative tests were developed based on the content on the content of the official pre-university course book taught in Iran. The subjects of this study were randomly assigned to one of the alternative assessments, and they received four formative tests and completed tasks based on the assumptions of the alternative assessment they belong to. The result suggest that language pra titioners can employ alternative assessments and co-operative learning, instead of focusing on traditional type of assessment, to promote learners involvement in their language learning processes.

Key Words: *Conference assessment, peer-assessment, self-assessment, washback*

Introduction

Assessment and testing are important dimensions of every teaching and learning practice. They provide teachers with the necessary quantitative information about their students' language ability and enable them to make sound professional judgments and decisions within the context of their classes. (Bachman, 1995) Brown and Hudson (1998) define the effect of testing, assessment, and assessment procedures on the language teaching curriculum as washback. Washback is also called back wash, test impact, measurement-driven instruction, curriculum alignment and test feedback in various places in the literature.

Although testing is considered an important part of teaching / learning activity, there seems to be some problems adherent to the significance of feedback and assessment procedures. In spite of the popularity of a variety of alternative assessments e.g. checklists, videotapes, audiotapes, self – assessment, peer-assessment and conferences (Brown and Hudson, 1998), teachers in generally resort only to traditional assessment based on their own evaluation. In most cases, scores provided by teachers do not make the students take the trouble of going through the material they have been tested upon to find the correct answers.

According to Bachman (1995, p. 56), "it is possible to improve learning and teaching through appropriate changes in the program, based on the feedback." Without feedback, there will be no reason to test, since there will be no decision to be made and therefore no information required. Moreover, the students involvement in the process of assessment has been proved to be pivotal to effective life – long learning and the development of professional competence (Brown et al., 1997) Most research findings support autonomous and collaborative learning as effective language learning procedures, and good language learners are now proved to be the ones who

take as much responsibility for their own learning as possible (Slavin, 1990, and Cotteral, 1995)

Considering the prime role of continuous feedback and integration of assessment, instruction and learning, this study investigates the washback effect of alternative assessment procedures including self-assessment, peer-assessment and conference on the achievement of male and female Pre-university learners. It also examines the potential effect of students being assessed by individuals who are at the same level of ability or by those who are at cross level ability. The definitions of these three types of alternative in assessment are presented below.

Conference

Conference assessments typically involve the students visiting the teacher's office, usually by appointment, to discuss a particular piece of work or learning process, or both. (Brown and Hudson, 1998) More importantly, conferences are different from other forms of assessment in that they focus directly on learning processes and strategies. (Genesee and Upshur, 1996)

Brown and Hudson (1998) state that teachers can use conferences to: (a) foster student reflection on their own learning processes; (b) help students develop better self-image; (c) elicit language performance on particular tasks, skills, or other language points; or (d) inform, observe, mold, and gather information about students. (p. 661)

Self-Assessment

Self-Assessment refers to the involvement of learners in making judgments about their own learning, particularly their achievements and the outcomes of their learning (Boud and Falchikov, 1990). it is a way for increasing the role of students as active participants in their own learning (Boud and Brew, 1995), and is mostly used for formative assessment in

order to foster reflection on one's own learning processes and results. (Dochy et al., 1999)

Peer- Assessment

Falchikov (1995) defines peer assessment as the process through which groups of individuals rate their peers. Brown and Hudson (1998) consider peer- assessment as "a variant of the self-assessment, which is similar to it except that, as implied by the label, students rate the language of their peers" (p.666). Topping et al. (2000) view peer-assessment as an arrangement for peers to consider the level, value, worth, quality or successfulness of the products or outcomes of learning of others in similar status.

Different positive effects are reported for peer-assessment. Orsmond et al. (1996) found that students did enjoy carrying out the peer-assessment fostered an appreciation for the individuals' performance within the group and interpersonal relationships in the classroom. Cheng and Warren (1997) conducted a research in the English Department of the Hong Kong University to gauge the students' attitudes prior to, and after a peer-assessment. Before and after the peer-assessment the students filled in a questionnaire with four items. The results of the questionnaire show that the students were mostly positive towards the peer-assessment. Williams (1992) found that the vast majority of students saw benefits in peer-assessment. Benefits were seen in three main categories: in comparison of approaches, in comparison of standards and in exchange of information. Sambell and McDowell (1997) studied six cases including peer-assessment, and found that students were generally positive towards an involvement in the assessment process. They mentioned that students were aware that peer-assessment had helped them to develop important skills, such as problem solving, improving self confidence, identification and belonging, and empathy with others.

Washback Effect and Language Testing

The relationship between testing and teaching has been the question of many educational practitioners. Testing and teaching have been alternatively seizing the priority, but they are so firmly interwoven that it seems very difficult to decide which one is leading the other. The influence of testing on teaching is called "washback". Alderson (1990) refers to washback as an issue, which is going to be one of the main concerns of language testers. Brown and Hudson (1998) define washback as the effect of testing, assessment and its procedures on the related language learning and teaching. Brown Hudson (2004, p.48) maintain that "a washback effect can lead to very negative consequences for a curriculum, or very positive ones- depending on what the effect is and how it is used by the curriculum developers". Messick (1996) defines washback as the extent to which the test influences language teachers and learners to do things "they would not necessarily otherwise do" (p.243).

More recently, a number of studies have confirmed the existence and complex nature of the washback effects (e.g., Alderson & Hamp-Lyons, 1996; Shohamy et al., 1996; Watanbe, 1996). All in all, the empirical studies to date seem to confirm the existence of the washback effect in various places with a variety of different effects. These studies are mostly concerned with the effects of tests on language learning and teaching, therefore a need is felt to investigate the comparative washback effect of some alternatives in assessment, namely self-assessment peer-assessment and conferences on learners' achievement. With this aim, this study intends to answer the following questions:

Research Questions:

Q1: Do self-assessment, peer-assessment and conference have any effect on Iranian EFL Pre-university learners achievement?

Q 2: Is there any relationship between sex variable and the impact of self – assessment, peer – assessment and conferences on students' achievement?

Q 3: Does the students' level of language proficiency in the peer-assessment groups have any impact on their achievement?

To examine the three research questions of this study, three null hypotheses were formulated:

Null Hypotheses:

Ho 1: Self – assessment, peer-assessment and conferences do not have any effect on Iranian Pre-university learners' achievement.

Ho 2: There is no relationship between sex variable and the impact of self – assessment, peer-assessment and conferences on students' achievement.

Ho 3: The students' level of language proficiency in the peer-assessment groups does not have any impact on their achievement.

Method

Subjects

The participants of this study were 220 Iranian male and female Pre-university pupils (110 males and 110 females) at the Pre-university Center in Zahedan. Both male and female subjects were divided into five groups each consisting of 22 male or female participants in separate classes. One pair group with 22 male and 22 female subjects were randomly selected as the control group, and the other four pair groups with a total of 176 students served as the four experimental group of the study receiving exactly the same four formative tests based on each lesson of their English textbook.

Instrumentation

The instruments used for data collection included a pre-test, four formative tests and a post-test. All tests contained 40 multiple choice items on structure (14), Vocabulary (14), Language function (2), cloze passage (5-6) and reading comprehension (3-4).

The reliability of all tests was computed through the KR-21 formula and the necessary modifications were made. It revealed to be 0.81 for the pre-test, 0.83 for the first formative test, 0.81 for the second one, 0.79, 0.84 for the third and the fourth one, and 0.81 for the post-test. Furthermore, an anterior examination of validity (Weir, 1990) was preferred in this study. Anterior validity mainly focuses on expert judgement on the construct of the items.

Procedure

The following steps were followed in the process of conducting the study: First, at the very beginning of the course, the pre-test based on the English textbook of third grade of high school was administered to all subjects participating in the study. The purpose was to ensure the homogeneity of the subjects with respect to their English language proficiency.

To check the homogeneity of the five groups at the outset of the experiment, a one – way ANOVA was performed.

Then, throughout the three-month semester of 24 sessions, the four experimental groups as well as the control group were given four formative tests, each based on the contents of the recently covered lessons of their textbook. The time of the administration of the tests was announced in advance.

After administering each test, the teachers collected the students' answer sheets and marked the blank spaces and the unanswered questions to assure accuracy in self- and peer-assessment groups.

In the self-assessment group, the students scored their own papers. They were asked to correct their wrong answers and bring the papers back the next session.

Within the two peer-assessment groups, the subjects of each class were divided into two levels of language proficiency (low and high) on the basis

of their scores on the pre-test. Those students who obtained or more were regarded as having a high proficiency and those who obtained less than 14 were viewed as having low proficiency. For the same – level peer-assessment group, students who were almost at the same level of language proficiency checked each other's papers. For the cross- level peer-assessment group, subjects who were at different levels of ability scored each other's tests. In order to be on the safe side and to avoid biased scoring the students' names were removed from the papers and their papers were coded by the teacher. After determining the students who had scored each other's papers, both members of each group were required to have sessions with each other and to work on problematic points.

For the fourth experimental group, the teacher him/herself marked the students' papers and the students had individual or group sessions with the teacher to discuss what they had done on the exams and to work on some diagnosed problematic points.

In all experimental groups, the received feedback provided the base for some remedial work and instruction.

In the control group, the traditional teacher assessment was employed for scoring the formative tests.

At the end of the course, all subjects took a post-test based on the contents of the pre-university textbook to investigate the effect of the treatments on their achievement during the period of instruction.

Results of the Pretest Stage

The descriptive statistics of the pretest is presented in Table 1. As presented in Table 1, the means of the five groups are almost the same. This is true concerning the standard deviation with an exception of the conference group. It indicates that the five groups are more or less homogeneous. However, a one-way analysis of variance was conducted to see if there was any significant difference between these groups. The F-ratio

[$F_{4,215}=.444$; $p = Ns$] shows that there is no significant difference in terms of the five groups' performance on the pretest at the beginning of the study. Thus, it can be safely concluded that the five groups participating in the study met the condition of homogeneity. Table 2 represents the results:

Results of the Post – Test Stage

After 3 month – 21 sessions -, the treatment was completely carried out and then an achievement post-test was administered at the end of the course. The descriptive statistics of the post-test is presented in Table 3.

For examining the first and the second null hypotheses of this research concerning the washback effect of the assessment procedures on learners' achievement and the relationship between sex variable and the impact of the aforementioned assessment procedures, a Two-way ANOVA was carried. Table 4 shows the relevant results.

As table 4 reveals, an F-ratio of 13.383 indicates a highly significant difference among the subjects in this study. In other words, the obtained results reveal that the first null hypothesis of the study is strongly rejected due to a significant difference among the groups. This can possibly be attributed to the effectiveness of the treatments.

Furthermore, the obtained results revealed no significant difference between the performance of subjects of different sexes in the post-test stage. Therefore, the second null was safely affirmed.

To find the location of differences, a post hoc scheffe test was performed, the results of which are presented in Table 5.

Table 5 reveals that there is a significant difference between the performance of the four experimental groups involved in the study and that of the control group (teacher – assessment group). Based on the mean differences between the control group and the four experimental groups reported in the first row of Table 5, the experimental groups can be ordered in the following way according to their improvement: (1) peer-assessment

(cross-level) groups (2) conference group (3) self – assessment group (4) peer – assessment (same – level) group. The control group revealed the least of all improvement that can be attributed to resorting to traditional teacher assessment and absence of feedback in this group.

The results reported in Table 5 also show that there is no significant difference between the two peer-assessment groups. In other words, those peers who were at different levels of language ability outperformed those peers who were at the same level of language ability, but the difference between the two groups was not significant. Therefore, the third null hypothesis of the research concerned with the impact of participants' level of language proficiency on their achievement in the two peer-assessment groups was safely affirmed.

Conclusion

The obtained data revealed that involvement of learners in the process of assessment led to a significantly better performance of the subjects. The better performance of the four experimental groups provides support for the promotion of learners' autonomy and collaboration are the two major premises. When students are engaged in assessing their own work, they think about what they have learned and how they learn. They are consequently more aware of their thinking and learning processes which encourage a deep, as opposed to a surface, approach to learning.

Furthermore, the out-performance of the peer-assessment (cross-level) group and the conference group supported the effectiveness of assistance and aid provided with a more advanced partner or a peer of higher status and power in terms of language ability. This can be due to the reciprocal process of giving and receiving feedback directly from the teacher or from a more proficient peer. While working with the teacher or a more advanced peer, students feel greater responsibility and this may have contributed to their better performance on the final achievement test.

The findings of this study are in agreement with Berger (1990), Zhang (1995), Orsmond and Marry (1996), Dochy et al. (1999), Mackenzie (2000), Topping et al. (2000) and Rasaei (2000) who have shown peer-assessment as having more favorable effects than self-and teacher assessment. However, the obtained results reject the findings of Hedgock and Lefkowitz (1992) and Caulk (1994) who found peer-assessment as having the same effect as that of teacher assessment.

Overall, it can be concluded that alternatives in assessment as a form of collaborative learning are beneficial because of the nature of interactions as well as introspection that occur through giving and receiving feedback. The conclusion of this research agrees with the contention that teachers can improve the effectiveness of instruction by resorting to formative tests while leaving the assessment of the tests to the students themselves or providing direct feedback and having discussion sessions with them concerning what they have done on the tests.

Pedagogical Implications

The major finding of this study was concerned with the out-performance of the four experimental groups in general, and the peer-assessment (cross – level) group and the conference group in particular. This recommends teachers to increase students' interaction through creating a cooperative atmosphere in their classes and allowing the students to take active roles in the assessment of the formative tests they take. The good performance of the conference group encourages students' engagement in giving and receiving feedback through providing the appropriate that will make the students reflect upon the processes of their learning, and ultimately improve their achievement.

Syllabus designers are also called for the specification of particular rooms in the course syllabi so that teachers can have enough systematically-planned

time at their disposal to administer the tests and gain and give the appropriate feedback.

References

- Alderson, J. C. (1990), Testing reading comprehension skills, *Journal of reading in Foreign Language*, 47/2, 265-287.
- Alderson, J. C.; Hamp Lyons, L. (1996), TOEFL Preparation Courses: A study of washback, *Language Testing*, 13, 280-297.
- Bachman, L. F. (1995), *Fundamental Considerations in Language Testin* , Oxford University Press.
- Berger, V. (1990), The effects of peer and self-feedback, *The CATESOL Journal*, November, 21-35.
- Boud, D. ; Brew, A. (1995), *Enhancing Learning through Self-assessment*, London, Kogan Page.
- Boud, D. ; Falchikov, N. (1990), Qualitative studies of self-assessment in higher education: a critical analysis of findings, *Higher Rducation*, 18, pp. 529-549.
- Brown, G.; Bull, J. ; Pedlebury, M. (1997), *Assessing Student Learning in Higher Education*, London: Routledge.
- Brown, J. D. ; Hudson, T. (2004), *Criterion-Referenced Language testing*, Cambridge: Cambridge University Press.
- Brown, J. D. ; Hudson, T. (1998), The alternatives in Language assessment, *TESOL Quarterly*, 32/4, 653-675.
- Castley, A. (1999), *Self – Assessment Forms*, University College Northampton.
- Caulk, N. (1994), Comparing teacher and student responses to written work, *TESOL Quarterly*, 28/1, 181-8.
- Cheng, W. ; Warren, M. (1997), Having second theories: student perceptions before and after a peer assessment exercise, *Students in higher Education*, 22, pp. 233-239.

- Cotteral, S. (1995), Developing a course strategy for learner autonomy, *ELT Journal*, 49/3, pp. 219-27.
- Dochy, F. Sergers, M. ; Sluijsmans, D. (1999), The use of self-, peer and co assessment in higher education: a review, *Studies in Higher Education*, Vol. 24, No. 3, pp. 331-347.
- Falchikov, N. (1995), Peer feedback marking: developing peer-assessment, *Innovations in Education and Training International*, 32, pp. 175-87.
- Genesee, F. ; Upshur, J. (1996), *Classroom-based Evaluation on Second Evaluation in second Language Education*, Cambridge: Cambridge.
- Hedgcock, J. ; Lefkowitz, N. (1992), Collaborative oral/aural revision in foreign language writing instruction, *Journal of Second Language Writing*, 1/3, pp. 255-76.
- Keaten, J. A. ; Richardson, M. E. (1999), A field investigation of peer assessment as part of the student group grading process, paper presented at the Western Speech Communication Association Convention, Albuquerque, NM.
- Longhurst, N.; Norton, L. S. (1997), Self-assessment in coursework essays, *Studies in Educational Evaluation*, 23, pp.319-330.
- Mackenzie, L. (2000), Occupational therapy students as peer assessment in Viva examination, *Assessment and Evaluation in Higher Education*, 25, pp. 135-148.
- Messick, S. (1996), Validity & washback in language testing, *Language Testing*, 13/3, pp. 241-256.
- O'Malley, J. M. (1999), Using authentic assessment in ESL classrooms, *Assessment & Evaluation in Higher Education*, Vol. 25, pp. 129-138.
- Orsmond, P. & Merry, S. (1996), The importance of marking criteria in the use of peer assessment, *Assessment & Evaluation in Higher Education*, 21, 99. 239-251.
- Rasaci, F. (2000), *Self-and peer-assessment Through Formative Tests*, MA Thesis, Tehran: Tehran University.

- Sambell, K. & Mc Dowell, L. (1997), The value of self-and peer-assessment to the developing lifelong learner, in: c. Rust (Ed.), *Improving student learning improving students as learners*, pp. 56-66. (Oxford, Oxford Center for staff and Learning Development)
- Shohamy, E. Dorista-Schmidt, S., Ferman, I. (1996), Test impact revisited: Washback effect overtime, *Language Testing*, 13, pp. 298-317.
- Slavin, R. (1990), Best-evidence synthesis: an alternative to mate-analysis and traditional reviews, *Educational Research*, 15/9, pp. 5-11.
- Topping, K. J.; Smith, E. F.; Swanson, I. ; Elliot, A. (2000), Formative peer assessment of academic writing between postgraduate students, *Assessment and Evaluation in Higher Education*, 25, pp. 149-170.
- Watanabe, Y. (1996b), Investigating washback in Japanese EFL classrooms: problems & methodology, *Australian Review of Applied Linguistics*, 13, 208-239.
- Weir, C. (1990), *communicative Language Testing*, Hemel Hempstead: Prentice Hall International.
- Williams, E. (1992), Student attitudes to wards approaches to learning and assessment, *Assessment and Evaluation in Higher Education*, 17, pp. 45-58.
- Zhang, S. (1995), Reexamining the effective advantages of peer feedback in the ESL writing class, *Journal of Second Language Writing*, 4/3, pp. 209-22.