

THE IMPACT OF TEST FORMS ON IRANIAN HIGH SCHOOL STUDENTS' KNOWLEDGE OF GENERAL ENGLISH

Zahra Bakhtiari¹ & Ramin Rahimy²

Department of English Language,
Tonekabon Branch, Islamic Azad University,
IRAN.

¹zahrabakhtiari13@yahoo.com, ²rahimy49@yahoo.com

ABSTRACT

The present study intended to investigate the impact of test forms on Iranian high school students' knowledge of General English. The question this study tried to answer was whether or not there was any difference among the means of the teacher - made test, school final exam and OPT test. To answer the question, 30 junior high school students participated in the three target tests: first, they were given an OPT test as the criterion test. Then, school final exam and teacher-made test were given to the students. Each of these tests were administrated after two weeks interval from the previous one. The contents of the three tests, in which vocabulary, grammar, reading comprehension and writing skills were examined, were the same except in the OPT test the content of writing section was different from the other tests. The data was analyzed through One- Way ANOVA. The results indicated that there was no significant difference among the means of the three tests.

Keywords: Test form, general knowledge, OPT test, final exam test, teacher-made test.

INTRODUCTION

The form of a test refers to its physical appearance. Considering the nature and varieties of attributes, language testers should utilize the most appropriate form of the test which would correspond to the nature of the attribute be measured. Therefore, to decide on the form of a test, certain parameters such as the nature of the attribute and the function of the test should be taken into account. Most assessments administered in school are criterion – referenced. Teachers create these tests based on the school's curriculum and learning expectations in a given subject area. Although personal experience is a good criterion, it is not a sufficient one for criticizing tests. Language tests play an important role in the investigation of effects of different instructional setting and techniques on language acquisition (Farhady *et al.*, 1994). Buck (1988) cited in Gholamhosseinzadeh (2012), believes that there is a nature tendency for both teachers and students to tailor their classroom activities to the demands of the test, especially when the test is very important to the future of the students, and pass rates are used as a measure of teacher success. Schwartz and Tiedeman (1957) cited in Frhady *et al.* (1994), stated that teacher – made tests are valuable because they measure students' progress based on the classroom activities, motivate students, provide an opportunity for the teacher to diagnose students' weaknesses concerning a given subject matter, and help the teacher make plans for remedial instruction, if needed. As Farhady *et al.* (1994) state since the function of a test influences its form and content, it should be determined in advanced. In order to determine the function of a test, three factors should be taken into account: characteristic of examinees, the specific purpose of the test and the scope of a test.

According to Heaton (1975) also added that a well-constructed classroom test will provide the students with an opportunity to show their ability to perform certain tasks in the language. Provided that details of their performance are given as soon as possible after the test, the students should be able to learn from their weaknesses. In this way a good test can be used as a valuable teaching device. According to Madsen (1993), testing is an important part of every teaching and learning experience. Well-made tests of English can help students in at least two ways. First of all, such tests can help create positive attitudes toward your class. In the interest of motivation and efficient instruction; teachers almost universally aim at providing positive classroom experiences for their students. He also noted that, a second way that English tests can benefit students is by helping them master the language. They are helped, of course, when they study for exams and again when exams are returned and discussed. Where several tests are given, learning can also be enhanced by students' growing awareness of your objectives and the areas of emphasis in the course.

Mehrens (1984) found that achievement testing served to broaden the curriculum; that is, testing encouraged teachers to add to, rather than replace existing instructional topics. Annandale, et al. (2003), states that effective teaching and learning stems from effective assessment and evaluation. According to Barry and King (2004), assessment relates to collecting, synthesizing and interpreting data about the knowledge and understanding, skills and attitudes of a person or group in order to facilitate decision-making.

BACKGROUND AND LITERATURE REVIEW

A brief review of the related literature would seem necessary to support the claims for the implementation of this study. According to Buck (1988) there is a nature tendency for both teachers and students to tailor their classroom activities to the demands of the test, especially when the test is very important to the future of the students, and pass rates are used as a measure of teacher success. According to Farhady et al. (1994) language tests play an important role in the investigation of effects of different instructional setting and techniques on language acquisition. Mehrens (1984) found that achievement testing served to broaden the curriculum; that is, testing encouraged teachers to add to, rather than replace existing instructional topics. AcSchwartz and Tiedeman (1957) cited in Farhady et al. (1994) state that teacher – made tests are valuable because they measure students' progress based on the classroom activities, motivate students, Provide an opportunity for the teacher to diagnose students' weaknesses concerning a given subject matter, and help the teacher make plans for remedial instruction, if needed. In spite of the usefulness of teacher – made tests, they have always been faced with students' complaints. These complaints have originated from the ambiguity of the content of the test and sometimes the irrelevance of such tests to instructional materials. Therefore, any test must be based on a pre- determined content to measure the students' knowledge at a given point of time. Such a test could be prepared by a teacher or a group of professional test – markers.

Recent studies raise questions about whether improvements in test score performance actually signal improvement in learning (Cannell, 1987; Linn, Grave & Sanders, 1989; Shepard, 1990). Other studies point to standardized tests' narrowness of content, their lack of match with curricula and instruction, their neglect of higher order thinking skills, and the limited relevance and meaningfulness of their multiple choice formats (Baker, 1989; Herman, 1989; Shepard, 1990). Several studies have investigated the ways in which testing influences teacher planning and instruction. Herman and Dorr-Bremme (1983) found relatively little influence of

standardized tests on teacher decision-making, for instance, in grouping students, planning instruction ,grading, diagnosing/prescribing compared to other sources of information available to teachers .In contrast, Salmon-Cox (1981) reported that teachers in Pittsburgh found the California Achievement Test to be useful in sequencing instruction, planning instruction and grouping students. Madaus (1988) noted that teachers taught to the test when they believed important decisions, such as student promotion, would be based on test scores. Corbett and Wilson (1988), in a study of Maryland schools, similarly found that schools redefined course objectives and resequenced course content in an attempt to improve test scores.

Annandale, et al. (2003) cited in Guy (2007), states that effective teaching and learning stems from effective assessment and evaluation. According to Bachman (1990), the fundamental use of testing in an educational program is to provide information for making decisions, that is, for evaluation. Barry and King (2004) cited in Guy (2007) believes that assessment relates to collecting, synthesizing and interpreting data about the knowledge and understanding, skills and attitudes of a person or group in order to facilitate decision-making. They go on to describe evaluation as the major step in the overall process pending assessment and that evaluation is when a value is placed on the worth of a score. According to A. Majid Hayati & Ehsan Askari, many language tests follow a psychological rather than linguistic theoretical framework, evidenced by the use of a single modality (such as a paper-and-pencil test that ignores spoken and oral comprehension) (Pray, 2005). Best and Kahn (1989) cited in Tasdemir (2010), assume that achievement tests attempt to measure what an individual learner has acquired. Achievement tests are particularly helpful in assessing individual or group status in academic learning. Achievement test scores are used to diagnose strengths and weaknesses and thus can be considered as a basis for awarding prizes, scholarship or degrees. They are also used in evaluating teachers, teaching methods and other factors which can be significant in educational situations. A language test which seeks to find out what candidates can do with language provided a focus for purposeful, everyday communication activities. Such a test will have a more useful effect on the learning of a particular language than a mechanical test of structure. In the past even good tests of grammar, translation or language manipulation had a negative and even harmful effect on teaching. A good communicative test of language, however, should have a much more positive effect on learning and teaching and should generally result in improved learning habits (Heaton, 1975).

According to Michele Vanpelt (2010), in the classroom, and within the school, a teacher will be utilizing many different types of assessments with his or her students. Some assessments may be formal, such as Standardized Achievement Tests or Informal Achievement Tests while others are less conventional. Both offer the teachers a good basis for determining what skills are being learned (or not learned) within the classroom. Test form in EFL setting has also received increasing attention over the past decade for improving language learning .According to Heaton (1975) ways of assessing performance in the four major skills may take the form of tests of: listening (auditory) comprehension, in which short utterances, dialogues, talks and lectures are given to the testees; speaking ability , usually in the form of an interview , a picture description, role play , and a problem –solving task involving pair work or group work ; reading comprehension , in which questions are set to test the students’ ability to understand the gist of a text and to extract key information on specific points in the text ;and writing ability, usually in the form of letters, reports , memos, messages, instructions, and accounts of past events ,etc. He also states that:

“...the evaluation of student performance for purposes of comparison or selection is only one of the functions of a test. Furthermore, as far as the practicing chief purpose of testing in schools. Although most teachers also wish to evaluate individual performance, the aim of the classroom test is different from that of the external examination. While the latter is generally concerned with evaluation for the purpose of selection, the classroom test is concerned with evaluation for the purpose of enabling teachers to increase their own effectiveness by making adjustments in their teaching to enabling certain groups of students or individuals in the class to benefit more. Too many teachers gear their teaching towards an ill-defined ‘average’ group without taking into account the abilities of those students in the class who are at either end of the scale. The test should also enable the teacher to ascertain which parts of the language programme have been found difficult by the class. In this way, the teacher can evaluate the effectiveness of the syllabus as well as the methods and materials he or she is using. The test results may indicate, for example, certain areas of the language syllabus which have not taken sufficient account of foreign learner difficulties or which, for some reason, have been glossed over.” (Heaton,1975)

THEORETICAL FRAMEWORK

The basic theoretical assumption behind the use of test in second language learning comes from the idea that language tests play an important role in the investigation of effects of different instructional setting and techniques on language acquisition. Baker (1989) cited in Glass & Stanely (1970), considers a test as a way of arriving at a meaningful decision. This means that the results of a test will lead us to the choice of a course of action also he believes that a test can be substitute for a more complicate procedure: a test is always a quicker or easier substitute for a more complicate decision-making procedure. According to Farhady (1982), in language testing, various theories, including those of discrete-point (Lado, 1961), integrative (Carroll, 1961), pragmatic (Oller, 1978), and functional testing (Upshur, 1979; Farhady, 1980) have been developed, and some have been supported by research results. These theories have generated numerous hypotheses involving such variables as the structure of language, instruments, test takers, and so forth. However, because of inadequate definitions, some of these hypotheses need to be reconsidered and the variables involved reexamined.

As Haynie (2003) states, the importance of testing in education and the many value-charged issues surrounding it make testing an important research topic. Research on testing has historically concerned standardized tests, while a large amount of evaluation in the schools is accomplished via teacher-made tests (Haynie, 1983, 1990a; Herman & Dorr-Bremme, 1982; Mehrens, 1987; Mehrens & Lehmann, 1987; Moore, 2001; Newman & Stallings, 1982; Stiggins, Conklin, & Bridgeford, 1986).

Heaton (1975) believes tests may be constructed primarily as devices to reinforce learning and to motivate the student or primarily as a means of assessing the student’s performance in the language. In the former case, the test is geared to the teaching that has taken place, whereas in the latter case the teaching is often geared largely to the test.

STATEMENT OF THE PROBLEM

Language testing is a complicated subject and much of this complication stems from problems of description and measurement which are particularly acute in linguistics and psychological investigation (Baker 1989, p.3).

Test – driven teaching and learning in Iran has a long history. The Iranian school final exam is a prerequisite as one of the criteria to determining whether the student is qualified or not to enter to upper class. Unfortunately, current English examination in Iran is heavily grammar- centered, which could have negative effect on learning. The examinee’s first impression of the test is its appearance. If the form of a test is harmonious with the nature of the attribute to be measured, it will attract the examinee (Farhady *et al.*, 1994). Yarmohammadi (1995) noted that the ultimate goal of EFL for a student in Iran is to master a foreign language and to reach for proficiency in all four language skills, namely listening, speaking, reading, and writing.

Empirical research is still lacking on the Iranian high school learners towards the effect of test form on their general English knowledge. With respect to this fact, test scores do not reflect real differences among learners. Therefore it reveals that most of the test- markers in Iran are not aware of the effect of test forms on learning. This study intends to determine the possible impacts of test forms on learning English as general knowledge and to reach true scores for evaluation of students’ learning ability.

Additionally, there are some concerns about some forms of testing being able to properly assess certain of the six skills. One of the most dreaded parts of school life has to be the class test. All the way through school, children have to take tests in one form or another. School tests take various forms - oral question and answer sessions, multiple choice questions, essay questions, practical demonstrations, and written short questions. These methods vary depending on the subject studied and the age of the students. Testing is extremely important however, because without it no teacher can really know how much the students have learned. This is necessary, not only in terms of the students but also for the teacher so that he or she can know where the class is holding when preparing the material for the next lessons. It can also show who the weaker and stronger students are - who needs extra help and who needs more of a challenge (Sachs, 2010).

The testing effect refers to the higher probability of recalling an item resulting from the act of retrieving the item from memory (testing) versus additional study trials of the item. However, in order for this effect to be demonstrated the test trials must have a medium to high retrieval success. Logically if the test trials are so difficult that no items are recalled or if the correct answers to the non-recalled items are not given to the test subject, then minimal or no learning will occur. In other words, test form has been found to be effective in improving different areas of a language (Bjork *et al.*, 1992). According to Farhady *et al.* (1994), language testers should utilize the most appropriate form of the test which would correspond to the nature of the attribute to be measured. Therefore, to decide on the form of a test, certain parameters such as the nature of the attribute and the function of the test should be taken into account. Most assessments administered in school are criterion – referenced. Teachers create these tests based on the school’s curriculum and learning expectations in a given subject area.

However, a survey in the literature reveals an absence of conclusive evidence of the positive effect of test form on improving learning ability in Iran. How test form would be more effective in learning ability of language learners has remained unanswered. Because of this gap, the

present study intends to investigate the influential effect of it on improvement of Iranian language learners.

RESEARCH QUESTION OF THE STUDY

The findings of the present study assisted the test – makers in designing the tests on the basis of the effectiveness of test form on Iranian High School Students' General English. Accordingly, the following research question was raised and guided this study:

RQ- Is there any difference among the means of teacher – made test, school final exam and OPT?

HYPOTHESES OF THE STUDY

In order to investigate the above mentioned research question, the following null hypothesis was proposed:

H₀ :There is no difference among the means of teacher -made test, school final exam and the OPT .

METHEDODOLOGY

Participants

Thirty junior high school students participated in this study. Because they were at the same level of education ,the subjects were homogeneous .They took a two - credit general English course. Sex were controlled and (all of the subjects were female). All of the participants took three tests with different forms but same content (except in writing section). This was done to investigate the effect of test forms on Iranian high school students' general English.

Materials and Procedure

To answer the research question relevant to this study a number of materials were devised. The materials used in current study were of three sorts and all of them were given as the pretest of the study. Since the study aimed at indicating the impact of test forms on Iranian high school students' general English , the forms of these three tests were different but the contents were the same (except in writing section) .They included the OPT test as criterion test, school final exam and teacher – made test. The three tests were used in this study consisted of the following sections: vocabulary, grammar, reading comprehension and writing. The participants were asked to answer each section. The reliability of teacher-made test estimated based on rater consistency ($r=0.76$). Each three tests included sixteen points for testing grammar, fourteen points for testing vocabulary, six points for testing reading and two points for testing writing skill. The total points for each of three tests were forty.

In order to investigate the impact of test forms on students' general English three tests in which the test forms were different were administered. All of the participants (N= 30) in experimental group (there was no control group) participated in these tests .The three tests were the OPT, teacher-made test and school final exam. The reliability of teacher-made test estimated based on rater consistency ($r=0.76$). There were two weeks intervals among administration of each test. The participants were asked to answer the following sections: vocabulary, grammar, reading comprehension and writing. The total points for each of three tests were forty. The time allowed for answering the whole sections was 90 minutes .The tests resembled in terms of time allocation

and their contents (except in writing skill). After administrating each test students' answer sheets were collected and were scored. Since all the tests had 40 points on the whole and the teacher decided to calculate the score of every student from 20, so the tests were divided by two. The scores in each test were calculated separately, and then the mean score in each test was calculated.

DATA ANALYSIS AND FINDINGS

The data obtained from testing the hypothesis of the study was analyzed via one-way ANOVA (Analysis of Covariance).

The Descriptive Analysis of the Data

This section focuses on the descriptive analysis of the obtained data in this study. Such analysis was done using the SPSS software. Table 1 shows the descriptive analysis of the three tests from the experimental group of the study:

Table 1. Descriptive analysis of the data

	<i>N</i>	<i>Mean</i>	<i>Std. Deviation</i>	<i>Std. Error</i>
OPT	30	11.5333	1.65536	.30223
final exam	30	10.7333	1.72073	.31416
teacher made	30	11.2000	1.78885	.32660
Total	90	11.1556	1.73472	.18286

According to table 1, the number of participants has been 30 in experimental group (N=30) and all of the participants participated in the present study (Missing Value=0.00). The mean and standard deviation of each of the three tests are presented in Table 1. Results showed that the means of each test was approximately close to each other. According to Table 1 the mean for the OPT test scores was 11.5333($x=11.5333$) as compared to the mean of final exam test scores which was 10.7333($x=10.7333$) and the mean of teacher-made test scores that was 11.2000($x=11.2000$). Although the means of OPT test was more than the means of the final exam and teacher-made tests but there was no significant difference between the means of them. Thus, it can be claimed that test forms had almost no effect on learning process and the difference between the means of three tests was not depend on test forms, maybe it happened by chance or any other reasons except test forms .

Inferential Analysis of the Data

In this section, the inferentially statistical results obtained from the experimental group were revealed. Such analysis was done via SPSS (Statistical Package for Social Science). Table 2 shows the inferential analysis of the data for the three tests in the experimental group of the study:

Table 2. The inferential analysis of the data (ANOVA F)

	<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>Sig.</i>
Between Groups	9.689	2	4.844	1.633	.201
Within Groups	258.133	87	2.967		
Total	267.822	89			

According to Table 2, although $F=1.633$ is more than one but it is not significant and noticeable. This finding revealed that forms had almost no effect on the quality of learning. Based on Table 2, it is shown that the $\text{sig}=0.201$ and since it is more than 0.05, the null hypothesis that “there is no difference among the means of teacher -made test, school final exam and the OPT” is accepted.

Results of Hypothesis Testing

The findings of the present study provided significant evidence responding to the research question which was “is there any difference among the means of teacher – made test, school final exam and OPT. Before analyzing the hypothesis, it will be repeated below:

H₀: There is no difference among the means of teacher – made test, school final exam and OPT.

According to the collected findings the null hypothesis of the study was supported. The results of the one way ANOVA of the study revealed that there is no significant difference among the means of teacher – made test, school final exam and OPT (see table 1) and as it was shown in table 2, there was no significant F ($F=1.633$) and $\text{sig}=0.201$ is more than 0.05 therefore it confirms the null hypothesis of the study with respect to this fact that tests forms have no noticeable effect on the quality of learning process and if it is not so ,there will be other reasons which affect learning .

DISCUSSION

General Discussion

The findings of the present study provided statistically significant evidence to find the answer to the research question of the study that targeted the possible existence of any difference among the means of teacher – made test, school final exam and OPT. Accordingly, findings of this study indicated that test forms did not show a significant impact on Iranian EFL learners’ performance in a test of general English.

Considering factors that influence test performance, the findings of this study contradict Bachman’s (1990) assertion that a test taker’s language performance is influenced by the characteristic of the method used to elicit the test takers’ language performance. Also, the findings of the current study are against Bachman and Palmer’s (1996) argumentation that characteristics of test methods always affect test scores to some extent and they further suggested that since it is impossible to eliminate the effects of test methods, it is necessary to control them as much as possible so that the tests will be appropriate for what they are used for.

Further, the findings of this study are in line with Moon, Brighton, & Callahan (2003) claim that standardized testing had little effect on teaching practices until the late 1970s despite the widespread use of testing.

SUGGESTIONS FOR FURTHER RESEARCH

In addition to the potential pedagogical benefits from this study, these findings also suggest a number of ideas for further research. Further researches can be carried out to investigate other skills such as listening, pronunciation to shed more lights on the possible effects of test forms. If there was a control group, the results were more accurate and acceptable. So, further researches can be designed with three groups: control group in which school final exam is given, one experimental group that participates in OPT test, and another experimental group with teacher-

made test. In addition to the potential pedagogical benefits from this study, these findings also suggest a number of ideas for further research.

In spite of the fact that the researcher asked participants to answer whole questions but most of them did not follow, especially in writing section in the OPT most of the participants missed this part or write very briefly. So it should be better if the future studies solve this problem and examine more writing criteria except word form errors. In this study the numbers of subjects were limited. It is hoped that other researchers working in this field would consider this point in their studies. It seems possible if future studies elicit information from a larger population, the result may be something different. Since the findings of the study may not be generalized to all students, it should be interesting if future studies are advised to expand the replications of this study to other language testing situation in Iran such as guidance schools ,another levels in high schools , universities or institutes . Furthermore, it can be a good topic for future research to study the effect of other standard tests such as Michigan in spite of OPT as criterion test.

REFERENCES

- Bachman, L.F. (1990). *Fundamental Considerations in Language Testing*. London: Oxford University Press.
- Bjork, R.A., Bjork, E.L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. http://bjorklab.psych.ucla.edu/pubs/RBjork_EBjork_1992.pdf
- FARHADY, H. (1982): "Measures of Language proficiency from the learner's perspective", *TESOL Quarterly*, 16, pp.: 43-59.
- Farhady, H. and Jafarpur, A. B. (2003). *Testing Language Skills From Theory to Practice*. Tehran: SAMT
- Gholamhosseinzadeh, S. (2012). The effect of Iranian school-leaving test of English on students learning of oral skills. *Modern Journal of Language Teaching Methods*. Retrieved from: <http://mjltm.com/index.php?option=content&t=co&i=60>
- Glass, G.V. and Stanely, J.C. (1970). Procedure and administration of the test. Retrieved from [http://www.thiqaruni.org/english/217/\(9\).doc](http://www.thiqaruni.org/english/217/(9).doc)
- Guy, K. (2007). Enhancing learning through assessment. Retrieved from: <http://www.helium.com/items/658044>.
- Haynie, W.J. (2003). Effects of Multiple-Choice and Matching Tests On Delayed Retention Learning In Postsecondary Metals Technology. North Carolina State University. Retrieved from: <http://scholar.lib.vt.edu/ejournals/JITE/v40n2/haynie.html>.
- Heaton, J.B. (1975). *Writing English Language Tests*. London: Longman
- Herman, J. L., and Golan, Sh. (1990). Effects of Standardized Testing on Teachers and Learning. Retrieved from <http://cse.ucla.edu/products/reports/Tech334.pdf>
- Hunter J. (2010). Reasons why final exams help education. Retrieved from <http://www.helium.com/debates/88031>.
- Madsen, H. (1993). *Techniques in Testing*. New York: Oxford University Press
- Motallebzadeh, K. Baghaee, M.P. (2011). Models of Language Proficiency: a Reflection on the Construct of Language Ability. *Iranian Journal of Language testing*, Vol. 1, No. 1, October 2011 ISSN 2251-7324
- Pray, L. (2005). How well do commonly used language instruments measure. Retrieved from: http://www.eric.ed.gov/ERIC_web_portal/recordDetail?accno=EJ742518
- Sachs, R. (2010). The importance of testing in school. Retrieved from <http://EzineArticles.Com/?expert=Ryan-sachs>
- Tasdemir, M. (2010). A comparison of multiple-choice tests and true-false tests used in evaluating student progress. *Journal of Instructional Psychology*, 37(3), 258-266.
- Van, P.M. (2010). Enhancing learning through assessment. Retrieved from: <http://www.helium.com/items/1763577>.