

## The Effect of Test Preparation on the Test Performance The Case of the IELTS and TOEFL iBT Reading Tests

Seyyed Ahmad Mousavi<sup>1</sup>, Ph.D. student in TEFL (corresponding author)

University of Tehran

Saleh Arizavi<sup>2</sup>, MA in TEFL

NamdarNamdari<sup>3</sup>

<sup>2,3</sup> Shahid Chamran University, Ahvaz, Iran

**Abstract:** *The main objective of this study was to see whether preparation influences the learners' reading comprehension of IELTS and TOEFL? Also as a subsidiary question, this study tried to find answer(s) to the following question: Does getting a high score imply enhancement in the reading comprehension ability? To this end, sixty EFL learners were selected to take part in a preparation program in which they were instructed different techniques and strategies to deal with the reading section of TOEFL and IELTS tests. Although the participants' scores enhanced in the post tests, the preparation program did differently on these two tests (IELTS and TOEFL). Generally, findings showed that the preparation effect was more visible on performance of the IELTS reading test than on that of the TOEFL reading test. The IELTS preparation group significantly outperformed the TOEFL preparation group on the IELTS reading test. In contrast, the TOEFL preparation group performed better than the IELTS preparation group on the TOEFL reading test, but this difference was not significant. These different effects of test preparation on test performance are clarified in the current study findings.*

**Key Words:** TOEFL / IELTS -preparation effects-reading ability-reading construct

### 1. Introduction

With the ever-increasing importance of IELTS and TOEFL tests for non-native speakers, as decisive gateways to better education opportunities and obtaining survival skills in the countries where English is spoken as first language, getting focused preparation for these instrumental tests has increasingly become a vital requirement. This special preparation may be of different sorts. For example, test familiarization is designed to ensure that prospective test takers are well versed in the general skills required for test taking and to help them gain familiarity with the procedures that are required to take a particular test. This type of preparation may entail, for instance, exposing test takers to the kinds of item formats they will encounter, making certain that they know when to guess, and helping them learn to apportion their time appropriately. Special preparation of this sort is generally regarded as desirable, as it presumably enables individuals to master the mechanics of test taking, thereby freeing them to focus on, and accurately

demonstrate, the skills and abilities that are being assessed (Powers, 2012). A variety of different types of test preparation: formal commercial coaching, school-offered test preparation programs, and test sponsor-provided test familiarization have been introduced (Powers, 2012).

Test developers, researchers, teachers and students/test takers have always been concerned with the issue of the effect of test preparation on the learners' performance of IELTS and TOEFL tests. However, on the other hand, from a more theoretical and pedagogical point of view, researchers and test designers often highlight the merit of a test which can independently measure the underlying ability of test-takers regardless of how well they have prepared or crammed for the test. In other words, if preparing results in a higher test score regardless of whether the candidate's ability has increased, thus the construct validity of the test is questionable (Nguyen, 2007). From a practical standpoint, however, teachers and students/test takers often feel the vitality of test preparation before being tested.

In Iran, preparation programs are normally deemed so much pertinent though there is hardly a clear and predefined program or plan for preparing the English language learners for special purposes like partaking in the TOEFL or IELTS exams. This problem stems, to a large extent, from the status of English learning and teaching in Iran, where English is learnt as a foreign language, so obviously there is little, if any, exposure to authentic situations to use real English as a back-up means to compensate for the typical deficiencies of classroom contexts and other testing events. Moreover, only the 'passive skills', i.e. reading and writing, (though this term has been harshly criticized by Chastain, 1988) are paid attention. In fact, reading skill is seen as the main source of providing learners with target language input, and expectedly less heed is given to reading section in preparation programs in Iran. The taken-for-granted expectation among teachers and test-takers is that the reading section of the international tests would be the trump card to get high score due to all the preparation learners received during their school years and college attendance. Given the importance attached to it, this skill was chosen as the independent variable for this study. Iranian learners often expect that if they attend TOEFL or IELTS preparation programs, then their scores on all skills will improve, especially on reading skill, for the reason mentioned earlier. Their expectations, however, are not always met.

The reasons for the gap between their expectations and what actually occurs have not been researched in-depth to date in Iran. It is thus argued that from theoretical and practical outlooks, it is useful to investigate the effect of test preparation on test scores in general, and on reading test scores in particular in an Iranian setting.

In this study, test candidates were instructed various combinations of test preparation materials to enable them to become more familiar with the reading section of the IELTS and TOEFL, with each of the item types they contain, and the strategies necessary to tackle this section in each test. The materials included full-length sample reading tests, and hints or tips for approaching each of these item types.

## 2. Review of related literature

### 2.1 The relationship between test preparation programs and test score improvement

Messick (1982, as cited in Powers, 2012) provided an insightful logical analysis of the ways in which special test preparation may impact validity, there appears to have been little empirical research to demonstrate how such practices may affect, for example, the relationship of test scores to other relevant measures. Most of the studies of test preparation have focused on the extent to which these practices cause spurious test score improvement. However, although relatively rare, researchers have also examined, in both a logical and an empirical manner, the effects of test preparation on the empirical relationships of test scores to other indicators of developed ability.

The impact of special preparation on test validity is a germane consideration. Though the assumption is sometimes made that preparation can serve only to adulterate the construct validity and impair the predictive power of a test, some kinds of special preparation may, by reducing irrelevant sources of test difficulty, actually improve both construct validity and predictive validity (Hayes & Read, 2004; Farnsworth, 2013).

A number of studies have investigated the relationship between test preparation programs and test performance scores (Anderman & Power, 1980; Bangert, Kulik, & Kulik, 1983; Powers, 1985, 1986, 2012; Geranpayeh, 1994; Bachman, Davidson, Ryan, & Choi, 1995; Thiel, 1995; Brown, 1998; Hayes & Watt, 1998; Celestine & Ming, 1999; Hayes & Read, 2004; Nguyen, 2007; Cho and Bridgeman, 2012; Sheshkelani, Ahari & Aidinlou 2012). The findings from this body of research, however, are to some extent contradictory: while a relationship between test preparation and test performance is acknowledged in some studies, doubts about this relationship have also been voiced. Thanks largely to Messick (1982), the question of the effectiveness of test preparation has been extended beyond the search for a simple dichotomous yes/no answer to the oversimplified question "Does preparation work?" Partly as a result, researchers now seem more inclined to examine the components of test preparation programs in order to ascertain the particular features that are implicated in its effectiveness. This complex picture can be seen in research conducted in both non-language and language testing contexts.

### 2.2 The effect of preparation programs on non-language tests

In research into non-language tests, incompatible findings can be seen regarding the effect of preparation programs. Anderman and Power (1980) studied the effect of a special preparation program on the verbal part of the Scholastic Aptitude Test (SAT). The researchers found that the special preparation program had very little impact on students' total scores in the SAT verbal section. Powers (1985) studied the effect of a special preparation program on scores of academic aptitude tests by using a representative sample (3%) from 5,107 candidates. His finding was

similar to that of Alderman and Powers (1980): the preparation program had diminutive influence on test-takers scores.

In contrast, Bangert, Kulik&Kulik(1983) used a “meta-analysis” or “the analysis of analyses” method to investigate the effectiveness of preparation programs on achievement test scores. By statistically analyzing the results of a large collection of individual studies, they concluded that preparation boosted achievement scores and that there was a positive correlation between the length of preparation period and the achievement score achieved. Similarly, Powers (1986) used a quantitative summary method to synthesize the results of ten previous studies on the effect of test practice on a number of test item characteristics. He found a strong relation between test preparation and (a) the length and complexity of the test instructions, and (b) the format of the test item.

Powers (1986) suggest that GRE analytical ability scores may relate more strongly to academic performance after special test preparation than under more standard conditions and that they may relate less to measures of other cognitive abilities (verbal and quantitative scores). No consistent effects were detected on either the internal consistency or the convergent validity of the analytical measure.

McLaughlin, Skaggs, and Patterson (2009) examined GED Test preparation activities and created eight mutually exclusive test preparation profile groups: public school adult education with or without a practice test, community college adult education with or without a practice test, individual study with or without a practice test, practice test only, and *none*. It was demonstrated that the group with the highest pass rates was the individual study with a practice test group, and the lowest pass rates were recorded for the public school adult education without a practice test group.

### **2.3 Research on the effect of preparation programs on language tests**

The effect of test preparation on language tests is a matter of some controversy;the effectiveness of preparation programs and products, unequal access to them, and concerns about the impact of preparation on the validity of test scores have always been at issue.

Bachman et al (1995) reported that test preparation did not produce a significant gain in test scores. Similarly, Celestine and Ming (1999) found that IELTS preparation did not make a significant difference to the scores of either average or high proficiency students from different disciplines. They went further by explaining that a preparation course did not have any effect on the test scores because ‘IELTS is a test of proficiency thus the knowledge of test-taking strategies cannot substitute for fluency’ (p. 46).

Another study referring to the effect of IELTS preparation on test performance in Asian students was carried out by Hayes and Read (2004). The findings revealed that though approximately half of the student population improved their scores, the difference was not significant. An earlier study by Hayes and Watt (1998) also focused on the effect of the IELTS

test preparation on the test performance of Asian students. They found that a two-month program with a testing rather than a teaching focus did not improve students' performance. A similar finding was reported by Elder & O'Loughlin (2003) in their study of IELTS score gains of 112 students after a three-month intensive English study in either Australian or New Zealand. Although the average score gain of these students was half a band overall with a slightly greater average improvement for listening, there was no advantage for the subset of students who had taken a course focusing specifically on test preparation. In a more recent study, Sheshkelani, Ahari and Aidinlou (2012) found that Iranian English major students without test preparation got lower scores in listening than those who were prepared for TOEFL.

However, a number of studies lend support to the effect of testing preparation programs on test performance. Brown (1998) compared the performance on an IELTS test of students in an IELTS preparation course and a more broadly focused EAP course. A positive effect of the IELTS preparation program on the students' performance was found. However, the sample sizes of the two groups in Brown's study were rather small (9 vs. 5) and only writing skills were researched. Thus the findings of his study need to be interpreted with caution. Brown called for 'a replication of the study with a larger population sample' (p. 36). Hayes and Read (2004) found a significant difference in the mean score in the IELTS listening sub-test of 12 Asian students after a course focused on IELTS preparation though their overall score improvement was not significant. In a larger scale study, Geranpayeh (1994) examined the comparability of TOEFL and IELTS scores across two groups (group A: 113 subjects and group B: 103 subjects). Group B had gone through the TOEFL preparation course and were more familiar with this test than group A. A test preparation effect was found as group B performed significantly better than group A.

Farnsworth (2013) states that most language assessment research on test coaching has looked at the effect of test formats on instructional practices, or so-called test washback, and not looked at score increases due to coaching practice. Existing research, however, has shown similarly modest results of coaching on scores (Hamp-Lyons (1991); Brown, 1998; Nguyen, 2007). This effect is intrinsically difficult to measure however, because genuine proficiency gains and score gains derived from mastering the test format and/or test-taking strategies are very hard to tease apart, and real proficiency gains likely occur alongside "test wiseness" score gains during test coaching instruction.

The issue that finds head from these studies on the effects of preparation programs is that researchers have been at odds in their findings regarding the effect of language test preparation on test performance. The differences in the findings might be due to many factors, as pointed out by Nguyen (2007), such as the kinds of tests and/or test formats, the sample sizes, the background disciplines, the background cultures, and the language proficiency levels of the studies' participants. In addition, the position of English, i.e., EFL or ESL is a critical factor, as perceived in Iranian context. Consequently, more research on this issue is needed to better understand the effect of test preparation on test performance.

In a similar vein, following Nguyen' (2007) study, which made comparison between the effect two preparation programs on learners' performance on the listening skill, this study aimed at investigating the effect of test preparation on the reading section of TOEFL iBT and IELTS and answering the following questions:

- 1: What is the effect of test preparation on test performance on the IELTS and TOEFL iBT reading section?
- 2: To what extent would the learners' performance in the reading section of IELTS and TOEFL iBT be indicative of their actual ability in reading skill?

### 3. Methodology

#### Research contexts and participants

Participants for the study were 60 Iranian students who were doing either the IELTS or TOEFL preparation courses in Ahvaz and Abadan cities, Khuzestan province in the southern part of Iran for the purpose of maximizing their chances of improving their reading score in TOEFL or IELTS or any similar tests like Ph.D. entrance examination in Iran. The IELTS preparation group consisted of 30 Ph.D. candidates majoring in different fields other than English at a private language institute; the TOEFL preparation group was also 30 Ph.D. candidates from different majors other than language attending another language institute.

#### 3.3 Instruments

It was not possible to obtain operational versions of either test because of security consideration; therefore, the test materials used in the study were the IELTS Sample reading test versions 2010 and the TOEFL iBT practice test online 2010. While this may place certain constraints on the validity of the study, it should be said that the IELTS Specimen practice test published by the British Council, IDP IELTS Australia, and University of Cambridge ESOL Examinations is closer to the real IELTS test than any other commercial IELTS practice tests available. Similarly, the TOEFL iBT practice test available in 2010 was closely contrived to the real TOEFL iBT test. The number of items for each sample test was 40 questions and the time allotted was 60 minutes for each test administration.

#### 3.4 Procedures

##### 3.4.1 Treatment

The IELTS preparation group was provided with instructions on IELTS 6 hours a week in which about 1.5 to 2 hours were devoted to reading skills and doing IELTS reading practice tests. The program lasted for 10 weeks. The English language proficiency of this group of students ranged from IELTS band 5 to IELTS 6.5 with the reading sub-band from 4.5 to 6.5. The TOEFL iBT group was introduced to the TOEFL preparation program for about 2 hours on the first day. After



that, instructions were focused on teaching reading skills in general and on the TOEFL iBT reading test in particular. For the participants in the TOEFL group, the training for the TOEFL iBT reading test ran for approximately 3 hours every other day and lasted for 2 weeks. The English language proficiency of the TOEFL preparation group ranged from 35 to 50 on the TOEFL iBT with the reading sub-band from 10 to 31. The time allocated for the preparation of the two groups and the level-off of their average scores were closely scrutinized so that the two groups would be as much the same as possible. This means that the outliers in both groups were excluded prior to study. Each group took the two tests in two different test administrations separately with an interval of 10 days to avoid any imitating of dissimilating effects.

### 3.4.2 Data collection

All participants in the study took both the IELTS and TOEFL practice tests. In order to avoid any possibility of a test practice effect, 30 participants took the *IELTS Specimen 2010 reading* test first and the other 30 took the iBT TOEFL reading test first. The data collection design can be summarized as follows (table 1).

Table 1: The data collection procedure in this study

IELTS preparation group		TOEFL preparation group	
Group 1	Group 2	Group 3	Group 4
15 learners	15 learners	15 learners	15 learners
Step 1: IELTS	TOEFL iBT	IELTS	TOEFL iBT
STEP 2: TOEFL iBT	IELTS	TOEFL iBT	IELTS

### 3.4.3 Data analysis

The effect of test preparation on test performance was investigated through a comparison of test performance between the two groups. T-tests were employed for the mean score comparison. It should be noted that the finding in this study ought to be interpreted with caution since the number of test takers is admittedly small for making generalizations, as maintained by (Camilli & Shepard, 1994; McNamara & Roever, 2006).

## 4. Results

### 4.1 The effect of preparation on the performance of IELTS reading test

The analysis of raw scores shows an effect of preparation of IELTS reading on the performance of the IELTS reading test for both groups of test takers. The descriptive analysis of the scores is provided in table 2 below.

Table 2: Descriptive analysis of the scores on the IELTS reading test

Groups	Maximum	Minimum	Mean	No.
IELTS Preparation	32	15	22.4	30
TOEFL iBT Preparation	24	13	18.7	30

As it is clear from table 2, the mean score and SD of the IELTS preparation group were higher than those of the TOEFL group. Particularly, the SD showed a greater homogeneity in the IELTS group. However, to examine whether there was a statistically significant difference between the two groups, an independent t-test was administered (table 3). The observed value (2.781) was higher than the critical t value ( $p=1.667$ ) at one *d.f.* and .05 level of significance. Therefore, we can conclude that there was a significant difference between the groups in the performance on the IELTS reading test.

Table 3: a comparison of means of the groups' performance on the IELTS reading Test

Groups	Mean score	Std. deviation	SE btwn. groups	$t_{obs}$
IELTS Preparation Group	22.4	4.901		
TOEFL iBT Preparation Group	18.7	5.428		
			1.331	2.781

#### 4.2 The effect of preparation on the performance of TOEFL reading test

Unlike the descriptive statistics obtained on the performance of the IELTS reading test, where there were clear differences between the two groups, there were less obvious differences in the TOEFL reading test scores across the two groups, as shown in table 4. The maximum and minimum scores of the two groups of test-takers are relatively close (Max. 28 vs. 32 and Min. 16 vs. 18). Besides, the mean score and SD of the two groups, though different, were also close.

Table 4: Descriptive analysis of the scores on the TOEFL reading test

Groups	Maximum	Minimum	Mean	No.
IELTS Preparation	28	16	23	30
TOEFL iBT Preparation	32	18	24.1	30

In order to find out if there was a statistically significant difference between the two groups, an independent t-test was used. The observed value was .419 which is smaller than the p-value (1.667) at one *d.f.* and .05 level of significance. Table 5 illustrates the inferential statistics. This led us to claim that there was no significant effect of preparation on the performance of TOEFL reading test between the two groups.

Table 5: a comparison of the means of the groups' performance on the TOEFL reading Test

Groups	Mean score	Std. deviation	SE btwn. Groups	$t_{obs}$
IELTS Preparation Group	23	3.383		



TOEFL iBT Preparation Group	24.1	3.483		
			2.620	0.419

## 5. Discussion and conclusion

The descriptive analyses of the raw scores, mean scores, and standard deviation of the two groups are suggestive of the effect of test preparation on test performance. This effect was particularly more evident in the IELTS reading test than in TOEFLiBT reading test. In short, it was found that there was a significant effect of IELTS reading preparation on test performance, while there was no such effect for the TOEFL reading preparation program. These findings support and contradict previous findings. On the effect of preparation on the IELTS reading, this study confirms Elder and O'Loughin (2003), Hayes and Read (2004), and Noguyn(2007), who found significant gains in the learners' performance on the reading section of IELTS test after an intensive IELTS preparation courses. However, this study contradicts what Celestine and Ming (1999) and Hayes and Watt (1998) voiced regarding the ineffectiveness of the IELTS preparation courses on learners' scores in the IELTS test.

Moreover, our findings are consistent with what Bachman et al. (1995) who claimed no effect of the TOEFL preparation on the learners' performance on the reading section. Likewise, in this study the results show no clear effect of the TOEFL iBT preparation on the learners' performance on the TOEFL iBT test. On the contrary, this study gainsays the findings reported by other researchers (e.g. Geranpayeh, 1994) interested in the same issue. Geranpayeh (1994) compared the effect of TOEFL preparation treatment with no preparation effect on the TOEFL test. He found that those who underwent TOEFL preparation outperformed those who received no instruction. Geranpayeh' study, however, is different from the present study in that it included an intact group to examine the effect of the independent variable, that is, the treatment. However, in the present study the effect of two different treatments, i.e. IELTS and TOEFL iBT preparation programs were investigated. Another striking difference was the TOEFL test employed in Geranpayeh's (1994) study (Noguyn, 2007). In that study, the test had less construct validity than its subsequent versions, since it was more 'amenable' to test preparation than the more recent ones (Noguyn, 2007). We believe the results obtained in the TOEFL iBT in this study were not affected by the test taking strategies presented in the treatment.

The differences in the effect of preparation programs on the learners' performances can be attributed to a number of considerations: firstly, the nature of the IELTS reading test which is quite different from that of the TOEFL reading section. The question types and the task types of the IELTS test impose different demands on the test taker to employ a variety of processes and authentic procedures to embark on answering the questions. On the other hand, in the TOEFL iBT there are a set of limited question types available that can be answered not necessarily by drawing on the strategies needed for the IELTS counterpart section. Therefore, it is tentative to claim that the questions in the IELTS are not only more difficult and demanding than the TOEFL

iBT ones, but also they tap on a wide range of askills and abilities beyond the straightforward reading strategies instructed in the classroom. Secondly, the challenge is to fold, and sometimes, unfold the combined strategies needed for the IELTS and without enough practice this is doomed to failure, while for the TOEFL one can compensate for the lack of language skill by relying on his/her global real-world knowledge. We also agree with Noguyen (2007) claim that combining different task types under one section would decrease the test method effect on the test performance, but it may have negative effect, instead. A third reason is the genre of the texts utilized in these two tests. The texts in the IELTS reading section are descriptive, discursive and academic-related subjects and thus can be regarded as richer than the TOEFL iBT texts, where the texts are restricted only to academic domain. This may lead us to conclude that the learners in the IELTS group are more familiar with a variety of genres and did better in the performance phase of the study, accordingly.

One important issue noted by Miyasaka (2000) and Rubenstein (2004) is how test preparation can potentially affect test validity. Miyasaka (2000) stated that the majority of large-scale assessment tests should be designed so that one could make reasonable inferences about the achievement levels of students with respect to content knowledge and/or skills within a given domain. The primary concern is whether a few months of preparation can significantly influence scores. In other words, if preparation does affect test scores, is the purpose of the test to measure the knowledge of the test-taker (pertaining to the topic of the given test)? If yes, how much knowledge is truly being measured, specifically if test-takers have access to tools that may artificially enhance their scores (Rubenstein, 2004, p. 398)?

ACT has noted that the earning of high scores on the ACT should not be merely a reflection of intrinsic talent or provisional preparation, but should reflect a level of accomplishment resulting from hard work, planning, and a solid commitment (ACT, 2005).

But the validity of the test score is compromised when test preparation artificially increases students' test scores without increasing mastery of the content domain, underlying subject-area knowledge, and/or testing skill (Perlman, 2004). Further, any test preparation practice that violates ethical standards would in effect nullify any validity of the students' test results (Miyasaka, 2000). Cho and Bridgeman (2012) express concern about the predictive validity of the TOEFL iBT expressed in terms of correlation with academic success. They maintain the students with higher TOEFL iBT scores tended to earn higher grade point averages (GPA) in their study and the TOEFL iBT provided information about the future academic performance of non-native English speaking students beyond that provided by other admissions tests. Combined with our results in this study, these observations can hardly lead us to conclude that a correlation might indicate a meaningful relationship between TOEFL iBT scores and preparation effect on test performance.

Contrary to what learners expected, their performance on the reading section of the two valid and accredited international tests was not backed up by their previous repertoire of skills acquired in their secondary and tertiary studies. It would be concluded that that formal education of English in our schools and universities doesn't provide our English students with necessary

and enough readiness for International and standard exams of English. So it is of high importance to have extra classes of skill teaching for English students as it is seen how effective and useful it was in for Non-English students. Last not the least is that this study was not a nation-wide, large scale and it is recommended that anyone willing to extend and make solid generalizations conduct it over a larger groups of learners with different levels of proficiency. The findings of the study will provide evidence to the score validity that concerns university admission officers and professors, i.e. whether accurate inferences and decisions can be made from students' TOEFL and IELTS scores to their actual English language proficiency.

Finally, ETS research on test preparation has been more than an academic exercise. It has resulted in significant—even dramatic—modifications to several tests that ETS offers. These changes are perhaps the clearest example of the impact of ETS's research on test preparation. However, there have, arguably, been more subtle effects as well. Now, when new assessments are being developed, the potential coachability of proposed new test item types is likely to be a factor in decisions about the final composition of a test. Considerations about test preparation figure into the *design* of tests, well before these tests are ever administered to test takers.

## 6. References

- ACT. (2005). *What kind of test preparation is best?* Iowa City, IA: ACT web flyer retrieved March 14, 2007, from [www.act.org](http://www.act.org).
- Alderman, D. L., & Powers, D. E. (1980). The Effects of Special preparation on SAT-Verbal Scores. *American Educational Research Journal* 17(2), 239-251.
- Bachman, L. F., Davidson, F., Ryan, K., & Choi, I. C. (1995). *An investigation of comparability of two tests of English as a foreign language*. Cambridge: Cambridge University Press.
- Bangert, R. L., Kulik, J. A., & Kulik, C. C. (1983). Effects of Coaching Program on Achievement Test Performance. *Review of Educational Research*, 53(4), 571-585.
- Brown, J. D. H. (1998). An Investigation into Approaches to IELTS Preparation, with particular focus on the Academic Writing Component of the Test. In S. Wood (Ed.), *IELTS Research Reports* (Vol. 1, pp. 20-37).
- Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. CA: Sage.
- Celestine, C., & Ming, C. S. (1999). The Effect of Background Disciplines on IELTS Scores. In R. Tulloh (Ed.), *IELTS Research Reports* (Vol. 2, pp. 36-51). Canberra: IELTS Australia Pty Limited.
- Chastain, K. (1988). *Developing Second-Language Skills: Theory and practice*. 3<sup>rd</sup> ed. Harcourt Brace Jovanovich, Publisher, Orlando, Florida: United State.
- Cho, Y., & Bridgeman, B. (2012). Relationship of TOEFL iBT® scores to academic performance: Some evidence from American universities. *Language Testing*, 29(3) 421–442. [ltj.sagepub.com](http://ltj.sagepub.com).
- Elder, C., & K. O'Loughlin. (2003). Investigating the relationship between intensive English language study and band score gain on IELTS. In R. Tolloh (Ed.), *IELTS Research Report* (Vol. 4, pp. 207-254). Canberra: IELTS Australia Pty Limited.

- Farnsworth,( 2013) Effects of Targeted Test Preparation on Scores of Two Tests of Oral English as a Second Language. *TESOL Quarterly* Vol. 47.1,148-186.
- Geranpayeh, A. (1994). Are score comparisons across language proficiency test batteries justified?: an IELTS - TOEFL comparability study. *Edinburgh working papers in applied linguistics* 5, 50 - 65.
- Hamp-Lyons, L.(1991). Assessing second language writing in academic context. Norwood,N.G:Albex
- Hayes, B., & Read, J. (2004). IELTS preparation in New Zealand: preparing students for academic module. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing.Research contexts and methods* (pp. 97-112). Mahwah: Lawrence Erlbaum Associates.
- Hayes, B., & Watt, L. (1998).An IELTS preparation course for Asian students; when practice doesn't make perfect.*EA Journal*, 16, 15-21.
- McLaughlin, Skaggs, and Patterson (2009).Preparation for and Performance on the GED Test. GED Testing Service®. Washington, DC:USA.
- McNamara, T., &Roever, C. (2006).Validity and the social dimension of language testing.*Language Learning*, 56(Supplementary 2), 9-42.
- Messick, S. (1982). Issues of effectiveness and equity in the coaching controversy: Implications for educational testing and practice. *Educational Psychologist*, 17, 67–91.
- Miyasaka, J. R. (2000, April). *A framework for evaluating the validity of test preparation practices*. Presentation in the symposium, instructionally corrupt test preparation: Can it be detected or deterred? Paper presented at the Annual Meeting of the American Educational Research Association, April 2000, New Orleans, LA.
- Nguyen,T.(2007)Effects of Test Preparation on Test Performance the case of the IELTS and TOEFL iBT Listening Tests.*Melbourne Papers inLanguage Testing*.1-23.
- Perlman, C. L. (2004). Practice tests and study guides: Do they help? Are they ethical? What is ethical test preparation practice? Chapter 27 in Wall, J. E., &Walz, G. R. (eds.). *Measuring Up: Assessment Issues for Teachers, Counselors, and Administrators*, 387-396. Greensboro, NC: CAPS Press.
- Powers, D. E. (1985).Effects of Coaching on GRE Aptitude Test Scores.*Journal of Educational Measurement*, 22(2), 121-136.
- Powers, D. E. (1986). Relation of Test Items Characteristics to Test Preparation/ Test Practice Effects: a Quantitative Summary. *Psychological Bulletin*, 100(1), 67-77.
- Powers,D.,E. (2012).Understanding the Impact of Special Preparation for Admissions Tests. ETS R&D Scientific and Policy Contributions Series.ETS SPC-12-02ETS Research Report No. RR-12-05.
- Rubenstein, J. (2004). Test preparation: What makes it effective? Chapter 28 in Wall, J. E., & Walz, G. R. (eds.). *Measuring Up: Assessment Issues for Teachers, Counselors, and Administrators*, 387-396. Greensboro, NC: CAPS Press.

Sheshkelani, A., Ahari.,H&Aidinlou,N. 2012). An Investigation of Relationship between Test Preparation and Test Performance of Iranian EFL learners on Listening Skill in TOEFL. International Journal of Basic & Applied Sciences IJBAS-IJENS ,12 , (2).

Thiel, T. (1995).*An Analysis of the Evolution of the IELTS and an Investigation of Its Validity*.University of Tasmania, Hobart.