An empirical study into the effects of Iranian test takers' personal attributes on their TOEFL scores: Reading, writing, and listening in focus

Amiryousefi, Mohammad ✉
University of Isfahan, Iran (M.Amiryousefi@Yahoo.com)

Tavakoli, Mansoor
University of Isfahan, Iran (Mr.Tavakoli14@gmail.com)

Abstract

TOEFL iBT is widely used as a certificating device, and is strongly claimed accountable by the people who utilize it. However, there are indications of the vulnerability of TOEFL iBT scores to the test takers' characteristics. The purpose of the present study is to provide empirical evidence to see if there is a relation between test anxiety, motivation and intelligence of test takers and their scores on TOEFL iBT reading, listening and writing sections. To carry out the present study, a Multiple Intelligences Development Assessment Scales (MIDAS) questionnaire, a Test Anxiety Scale, a Motivation Test Battery (MTB) and a survey question detecting test anxiety provoking factors were given to 60 subjects attending TOEFL iBT classes at ACECR. Then the relation between their scores on TOEFL iBT test and their responses to the above-mentioned questionnaires were measured. The results revealed that first, there was a relation between test takers' scores on TOEFL iBT writing section and test anxiety and kinesthetic intelligence, and between listening and musical intelligence and second, some factors like time limitation and test length provoked test anxiety.

Keywords: TOEFL iBT; test anxiety; test anxiety provoking factors; multiple intelligences; motivation

An empirical study into the effects of Iranian test takers' personal attributes on their TOEFL scores: Reading, writing, and listening in focus

1. Introduction

Language testing is mainly concerned with the extent to which the obtained scores truly mirror test takers' underlying abilities in a particular area in a specific testing situation. It, therefore, deems necessary to answer some questions before any attempts are made to devise a test. These questions are: What is it that we are testing? What tasks and items can elicit it? In what format should the tasks and items be presented to the test takers? In what mix should the items be included to represent the ability or the construct appropriately? What scoring procedure should be taken to be able to extrapolate the test performance to the underlying behavior (Weir, 2005; Fulcher & Davidson, 2007)? Throughout the past decade, test developers and experts have devoted much of their interest and energy to develop a theoretical view of language ability to investigate the nature of language proficiency, and to develop and apply more sophisticated statistical tools to analyze language tests and test takers' performance to provide sophisticated answers to these questions (Bachman, 1990; Bachman & Palmer, 1996).

However, research in language testing has indicated that language ability is not the only thing that affects test takers' performance on a language test. Test performance can be affected by a wide range of factors, which can counteract with test takers' language ability and affect their test performance positively or negatively (Lumley & O'Sullivan, 2005). These construct-irrelevant factors, as Messick (1996) puts it, are considered as the potential sources of test bias, which can make the obtained scores unrepresentative of the underlying ability that a language test wants to measure, and put the whole testing process at stake (Takala & Kaftandjieva, 2000; Messick, 1996).These factors are classified by Bachman (1990) into three groups (a) test method facets, which make the how of language testing and are considered as factors that provide the contextual features for language performance on language tests; (b) personal attributes like cognitive and affective characteristics, real-world knowledge, sex, age, motivation, anxiety, ambiguity tolerance, etc. that are irrelevant to test takers' language ability; and (c) random factors like mental alertness, emotional state, test environment, etc. which are unpredictable and temporary.

Test performance is, therefore, a function of a test taker's language ability and the effects of one or more of these factors. The examination of such factors and the investigation of their effects have long been recognized essential in the estimation of both reliability and validity (Bachman, 1990; Skehan, 1989, 1998; Tavakoli, 2009). Taking into account these considerations and the status of TOEFL iBT as a proficiency test around the globe, the purpose of this study is to measure the probable impacts of test anxiety, multiple intelligences and motivation on listening, reading and writing scores of this test.

2. Theoretical background

Test fairness is an aspect of language testing, which is hotly debated in the literature. The central aim of test fairness discussions is to make tests free from bias and to contribute to testing equality (Kunnan, 2007). A test may be considered biased when students having the same language ability perform differently. Different factors like content or language of the test when the content of the test is biased to test takers from certain groups, religions, ethnicities etc., standards of the test that may affect test takers from different groups differently, and test takers' personal attributes such as age, learning strategies and styles, attitude and motivation, aptitude, intelligence and anxiety can contribute to test bias (Kunnan, 2004, 2007; Weir, 2005; Brown, 2004).

Test takers' personal attributes can, therefore, affect their performance on a given test in one way or another, and test takers with specific predispositions may do better or even worse on a test if the test is biased to and has a

direction, whether positive or negative, toward their predispositions (Bachman, 1990). Hence, these features can affect the validity of a test and can cause measurement errors. A major concern in the design and development of language tests is, therefore, to minimize the effects of these factors that are not part of the language ability (Bachman & Palmer, 1996; Messick, 1989, 1996). Test anxiety, multiple intelligences and motivation are amongst these features.

2.1 Test anxiety

Test anxiety is a form of anxiety concerning apprehension over academic evaluation, which comes from a fear of failure (Horwitz & Young, 1991). It is believed to be caused by certain factors like time limit, test techniques, test format, test length, testing environment and ambiguity of test instructions (Young, 1999). Hsu (2012) also believes that the present condition of the world has made the graduates enhance their English proficiency and strive to get better grades which can lead to test anxiety among them. Some researchers believe that test anxiety negatively affects language-learning process and the test outcomes. It is believed to cause cognitive interference while preparing for examination, taking examination or both. Test anxiety during the preparation for a test may lead to poor understanding and organization of the concepts and hence makes the retrieval cumbersome. While test anxiety during the test decreases test takers' attention and thereby increases the number of errors (Ohata, 2005; Cassady & Johnson, 2002; Sarason, 1980, 1986). It seems that test anxiety inversely affects test performance. Chapell et al. (2005), however, believe that for some students, some anxiety might be a good thing. It might motivate longer periods of study and more careful attention to questions on the exam.

2.2 Multiple Intelligences

One of the most compelling, yet controversial new approaches to education reform is Multiple Intelligences Theory, or MI. It was first conceived of by Howard Gardner in 1983. He challenged the commonly held view defining intelligence as a general factor and believed that intelligence is too narrowly defined in this way. Gardner was seriously against testing intelligence by making the person to do isolated tasks that he has never done and may never choose to do. Rather he believed that intelligence as an ability of problem solving must be tested in a natural context (Armstrong, 2009). Multiple intelligences theory, in a nutshell, is a pluralized way of understanding the intellect. Recent advances in cognitive science, developmental psychology and neuroscience suggest that each person's level of intelligence, as it has been traditionally considered, is actually made up of autonomous faculties that can work individually or in concert with other faculties. Gardner originally identified seven such faculties, which he labelled as "intelligences": Musical Intelligence; Bodily-Kinesthetic Intelligence; Logical-Mathematical Intelligence; Linguistic Intelligence; Spatial Intelligence; Interpersonal Intelligence and Intrapersonal Intelligence. Recently, he added an eighth intelligence to the list: the Naturalist Intelligence and has discussed for a ninth one (Visser, Ashton, & Vernon, 2006). Development of each type of aforementioned mentioned intelligences is an interaction of the following factors:

1. biological endowment—including hereditary or genetic factors and insults or injuries to the brain before, during, and after birth;

2. personal life history—including experiences with parents, teachers, peers, friends, and others who awaken intelligences, keep them from developing, or actively repress them;

3. cultural and historical background—including the time and place in which you were born and raised and the nature and state of cultural or historical developments in different domains (Armstrong, 2009, p.27).

MI theory has opened the door to new and wide range of teaching strategies, materials, curriculum development and lesson planning. In MI era, a successful teacher is a person who first recognizes his own bases of intelligences and looks for areas waiting for further development. Otherwise, teachers, for example, without a

base of spatial intelligence will avoid using visual and art activities, and this type of intelligence will not be taped in their classes. Instead, they need to shift continually from an intelligences to another and use a combination of all types of intelligences in lesson planning and material development in a way that all students can have their strongest intelligences addressed at least some of the time (Armstrong, 2009).

Loori (2005) believes that preferences of intelligences are at work in educational settings. The results of his study indicates that males are more successful in tasks that involve logical and mathematical intelligences, whereas females are more successful in tasks involving intrapersonal intelligence. Intelligence type can, therefore, affect learning and testing. Mahdavy (2008), for example, believes that linguistic intelligence significantly contributes to the listening proficiency and Richards and Rodgers (2001) believe that musical intelligence contributes to English pronunciation. Razmjoo (2008) , however, found no significant relationship between language success and the types of intelligences in particular.

2.3 Motivation

Motivation is defined as a derive, a desire or an emotion which has three major components namely motivational intensity, a desire to learn, and a positive attitude which fuels L2 learners to strive to learn (Gardner, 1985; Brown, 1994: Purpora, 2004). According to Gardner, there are two types of motivation namely integrative and instrumental. An integratively motivated L2 learner shows interest in learning about the culture and the people of the target language. Integrativeness is a desire to learn a language in order to "come closer to the other language community" (Gardner, 2001, p. 5). Integratively motivated learners are believed to invest more effort on the learning process and enjoy it more (Lamb, 2004).An instrumentally motivated learner, on the other hand, has more pragmatic considerations in his or her mind regarding L2 learning, such as obtaining a job or earning more money. An instrumentally motivated learner approaches language learning for utilitarian objectives (Tavakoli, 2009). There are different views about the possible relationship between testing and motivation. The Assessment Reform Group (2002) believes that testing provides incentives for the students and test takers to improve their performance. They also believe that tests can be motivating for those that anticipate success and can increase the gap between high and low achieving test takers. They, however, believe that motivation has some components (Fig.1) and assessment is only one of them. Therefore, the effects of testing on motivation may be small but the effects of motivation on testing can be considerable.
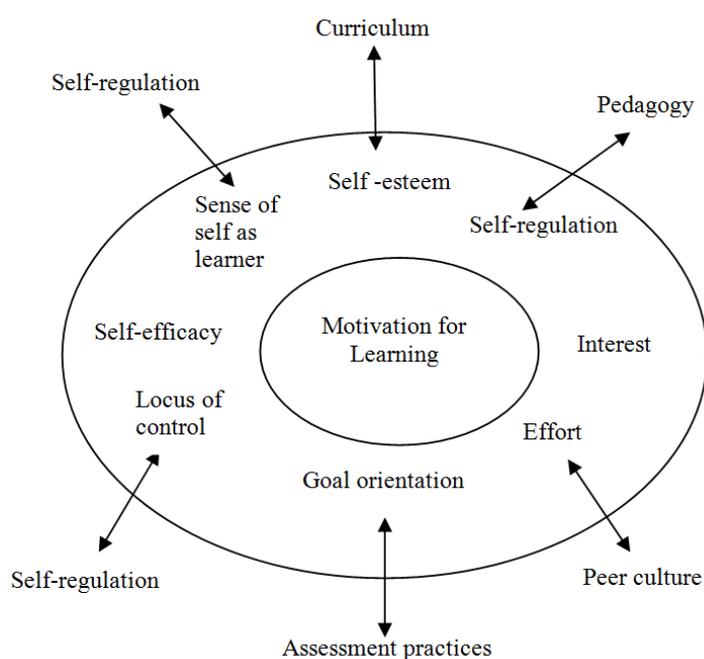

Figure.1. The relation and the components of motivation for learning

The Assessment Reform Group defines self-esteem as how one values oneself as a person and as a learner, self-efficacy as how capable one feels of succeeding in a learning task, self-regulation as the capacity to evaluate one's own work and to make choices about what to do next, goal-orientation as whether one's goal is to learn in order to understand or to perform well on a test (which may not reflect secure learning), interest as the pleasure from and engagement with learning, effort as how much one is prepared to try and persevere, locus of control as how much one feels in control of learning as opposed to being directed by others, and sense of self as a learner as how confident one feels of being able to learn from the classroom experiences provided (Assessment Reform Group ,2002, p. 3).

Motivation is, therefore, studied as a factor interacting with test takers' performance and affecting test results. Researchers are now concerned with finding what kind of motivation leads to purposeful engagement of test takers' abilities on a language test. Some researchers, however, claim that integrative motivation is more advantageous (Lamb, 2004; Kunan, 1995).

2.4 The purpose of the study

Taking the effects of factors such as anxiety, motivation and MI on test performance and the status of TOEFL iBT around the world, the purpose of this study was to seek the answers to the following questions:

➢   Q1: Is there a relation between test takers' scores on TOEFL iBT listening, reading and writing tests and test anxiety, MI and motivation?

➢   Q2: What factors provokes test anxiety among TOEFL iBT candidates?

3.   The study

3.1 Participants

Sixty students (28 males and 32females) attending TOEFL iBT preparation classes at Iranian Academic Centre for Education, Culture, and Research (ACECR) in Fall 2013 participated in the study. Their age ranged from 20 to 36. They had learned English for at least two years before they attended the TOEFL classes. They were tested at the beginning of the TOEFL iBT preparation course to neutralize the effects of the coaching. They were all at the intermediate level based on the results obtained from both the Oxford Placement Test (OPT) (2004) and TOEFL iBT (2010).

3.2 Instruments

To carry out the study several instruments were used to elicit the required data. To assess the participants' intelligence types, Multiple Intelligences Development Assessment Scales (MIDAS) (Tirri & Nokelainen, 2011) was used. This instrument contained 30 response categories classified under eight main themes. Each theme represented one type of intelligence as classified by Gardner (1985). Students rated each category on a 5-point Likert scale with anchor points of always: 1 to never: 5. Test Anxiety Scale (TAS) (Sarason, 1984) was also used to assess the participants' test anxiety. This test consisted of 24 response categories which were rated on a 5-point Likert scale type ranging from always: 1 to never: 5. Gardner's Motivation Test Battery (MTB) (Gardner, 1985) was also given to the participants to determine their motivation types. This instrument contained eight response categories adapted to a 5-point scale, ranging from strongly agree: 1 to strongly disagree: 5.

Finally, a survey question was used to detect the factors that provoked test anxiety among the participants. This question asked the participants to name factors that provoked test anxiety and arrange them from the most important to the least important. This question was developed by the researchers and then reviewed by three experts to ensure its content validity. The reliability estimate of the instruments used is shown in Table1.

Table 1

The reliability statistics of the instruments used in the study

	Cronbach's alpha	N of Items
MIDAS	0.91	30
TAS	0.89	24
MTB	0.82	8

Cronbach's alpha is bigger than 0.7 ($\alpha > 0.7$) for MIDAS, TAS and MTB which indicates the reliability of these instruments. These instruments have also been used in the same format by many researchers in the literature and have shown acceptable indexes of reliability and validity (Tirri & Nokelainen, 2011; Sarason, 1984; Gardner, 1985).

3.3 Procedures

At first the Oxford Placement Test (2004) was given to 80 students and 60 students who, based on the results of OPT, had the same level of proficiency were selected. A week later a TOEFL iBT test released by ETS in 2010 was given to them. After the TOEFL test, the participants were given a short break, and then MIDAS, TAS, MTB and the survey question described earlier were given to them. The performance of the test takers on the TOEFL test was scored on the basis of the scoring procedures used by ETS for TOEFL iBT test**.** For the writing section, however, three experienced raters were recruited to rate the writing tasks. They were all experienced English teachers and were already familiar with TOEFL scoring procedures. To obtain inter-reliability, an average of the three raters' scores was used to arrive at the final score. For MIDAS, TAS, and MTB the median of the participants 'responses were, however, calculated and analyzed. The participants' responses to the survey question were also read and categorized.

4.   Results

At first, three raters who were already familiar with the scoring procedures of TOEFL iBT test were recruited. They were knowledgeable and experienced teachers who had taught TOEFL preparation courses for years. The procedures were reviewed in an introductory session and then the participants' performance on writing, reading and listening parts of TOEFL iBT was scored. Their responses to MIDAS, Test Anxiety Scale, and MTB were collected using the procedures described earlier and fed into SPSS version 16 to measure the strength of the relation existing among them. The results obtained are described in the following parts. Table2 shows the descriptive statistics of the participants' scores on writing, reading and listening parts of TOEFL iBT. As shown, the general average for the listening part is lower than the average for the other parts. It may mean that they found the listening part a little more difficult.

Table 2

Descriptive statistics of the scores obtained

		writing	reading	listening	Total	age
N	Valid	60	60	60	60	29
	Missing	0	0	0	0	1
Mean		17.27	17.70	15.23	16.7222	25.69
Std. Deviation		2.690	4.268	2.861	2.10288	3.219
Minimum		13	10	10	13.33	20
Maximum		23	26	24	23.00	35

To ensure sample homogeneity, Kolmogorov-Smirnov Test was used. In Table3, the words reading, writing and listening represent the participants' scores on these parts of TOEFL iBT test, the word total shows their total score on all these three parts, the word test anxiety shows their responses to TAS, the words musical, kinesthetic,

spatial, linguistic, interpersonal, intrapersonal, naturalist, mathematic and logical represent their responses to the response categories existing in MIDAS for each of these intelligence types, and the word motivation shows their responses to the response categories in MTB.

As shown in Table 3, for reading, the total score, musical, spatial and the motivation scores the significant value is smaller than 0.05 ($\alpha = 0.05 <$ Sig = 0.562) which shows that sample homogeneity does not exist for these parts. A non-parametric test like Pearson Correlation Test must, therefore, be used to explore the relationship between the participants' scores on the reading, listening and writing parts of the TOEFL iBT test and their responses to MIDAS, TAS, and MTB.

Table 3

The results of Kolmogorov-Smirnov test

	Age	Reading	Writing	Listening	Total	Test Anxiety	Musical	Kinesthetic	Spatial	Linguistic	Interpersonal	Intrapersonal	Naturalist	Logical Mathematical	Motivation
Stat	.170	.167	.146	.160	.164	.104	.180	.128	.199	.139	.126	.116	.105	.080	.199
Sig.	.032	.037	.115	.055	.045	.200	.017	.200	.005	.160	.200	.200	.200	.200	.005

As indicated in Table 4, at 0.01 levels, there is a correlation between writing and test anxiety and a correlation between listening and kinesthetic, and at 0.05 levels, there is a correlation between writing and musical intelligence. Pearson's *r* is, however, 0.653 for the relation between writing and test anxiety which is closer to 1 indicating a rather stronger relation between these two.

Table4

The results obtained from Pearson Correlation test

		Test Anxiety	Musical	Kinesthetic	Spatial	Linguistic	Interpersonal	Naturalist	Logical Mathematical	Intrapersonal	Motivation
writing	*R*	.653**	.404*	-.015	.047	.102	.146	.066	.117	-.019	.186
	Sig.	.005	.027	.938	.805	.592	.441	.731	.537	.922	.324
	n	60	60	60	60	60	60	60	60	60	60
reading	*R*	.042	-.078	.038	.090	.069	.132	.043	.211	.118	.033
	Sig.	.826	.683	.840	.636	.717	.488	.823	.263	.535	.864
	n	60	60	60	60	60	60	60	60	60	60
listening	*R*	-.065	.216	.484**	.388*	.245	.228	.153	.329	.333	.049
	Sig.	.735	.251	.007	.034	.191	.225	.420	.076	.072	.798
	n	60	60	60	60	60	60	60	60	60	60
total	*R*	.319	.143	.102	.277	.081	.123	.066	.259	.148	-.003
	Sig.	.920	.449	.593	.139	.670	.517	.728	.167	.434	.986
	n	60	60	60	60	60	60	60	60	60	60

Note. *. Correlation is significant at the 0.05 level (2-tailed). **. Correlation is significant at the 0.01 level (2-tailed).

It can, therefore, be concluded that the participants with kinaesthetic intelligence performed better on the listening part, and those with musical intelligence performed better on the writing part of TOEFL iBT test. The participants who had some degrees of test anxiety also handled the writing part better than the others.

The participants' responses to the survey question which asked for the factors that provoked test anxiety among them were also categorized into 3 factors and shown in Table 5 based on the level of importance given to them by the participants from the most important: 1 to the least important: 3. The time limit which is usually set for exams like TOEFL caused the most amount of stress among the participants. Also, they had problems with the length of the test and they believed that the test was lengthy and they were afraid that they might not be able to finish it on time. Some participants, however, criticized their lack of enough self-confidence and considered it as the source their anxiety.

Table 5

Test anxiety provoking factors

Importance	Factors
1	Time limit
2	Test length
3	Lack of enough self confidence

5. Discussion

TOEFL iBT is being widely used in the world to measure English proficiency of the candidates and its scores are the basis of important decisions like acceptance or non-acceptance of specific candidates for work or education. It may, therefore, have significant consequences on the lives of the people who take it. Test fairness is also put forth by scholars in language testing (Kunnan, 2004, 2007; Weir, 2005; Brown, 2004) to discuss the effects of construct-irrelevant factors such as test takers' personal attributes like anxiety, motivation and multiple-intelligences on their performance on a test. Theses scholars believe that all tests specifically high-impact tests like TOEFL iBT must not be biased toward specific groups of candidates. They rather must provide equal conditions for all.

This study was, therefore, an effort to empirically explore the relation between test anxiety, MI and motivation and the performance of the test takers on listening, reading and writing sections of TOEFL iBT test. The findings of the study revealed that there was a positive relation between test anxiety and the writing performance, and kinesthetic intelligence and the writing performance of the participants, and also between musical intelligence and their listening performance. The answers of the test takers to the question exploring the possible causes of test anxiety ranked time limit, length of the test and lack of enough self-confidence among the most common test anxiety-provoking factors.

Writing is a difficult task for some language learners. These learners may not have any special difficulties with speaking which is also another productive skill, but they somehow become uncomfortable when it comes to writing. These learners believe that writing is time-consuming and they think structures, styles, words, etc. must be selected carefully to create an acceptable piece of writing. Writing, therefore, demands a good mastery over these areas which sometimes makes test takers anxious. Test anxiety, on the other hand, can make language learners perfectionist. "*Test-anxious students often put unrealistic demands on themselves and feel that anything less than a perfect test performance is a failure*"(Mahmoodzadeh , 2013, p. 64)**.** It means that these learners are more self-conscious and very sensitive toward their mistakes which can make them attempt to do well on the tests to get better grades (Gregersen & Horwitz, 2002). The positive relation between test anxiety and the writing performance of the participants in this study can, therefore, be attributed to the fact that the higher the level of test anxiety, the more conscious the participants are toward their mistakes and, therefore, the better their writing will be. Brown (2004) also believes that a little stress about a task at hand is going to be facilitative and being so soft on the students may have a debilitative effect on them, because they may be carefree and have no anxiety or concentration. Test anxiety can, however, be debilitative if its level is too high (Rollinson, 2005).

People with kinesthetic intelligence also love figuring out how things work. These people have high creativity and they like to build new things. This sort of intelligence is something that writers usually have (Saleem, 2008). The relation between kinesthetic intelligence and the writing performance of the participants can, therefore, be attributed to the fact that the subjects with this type of intelligence can figure out the whole writing task in advance, and use their creativity to improve the structure and the quality of their performance.

Fonseca-Mora, Toscano-Fuentes, and Wermke (2011) in their study to examine the relation between musical intelligence and language aptitude concluded that people with musical intelligence can better perceive rhythm, melody and crucial elements of language which can lead to better and faster language learning and

understanding. The correlation between musical intelligence and the participants' listening performance can be interpreted in the light of Fonseca-Mora et al.'s conclusion that learners with this type of intelligence can perceive and understand the crucial elements of a listening task better and hence do better on it.

This study, therefore, further proves the heterogeneity nature of test takers and the complexities and difficulties involved in accounting for them in test tasks and items. It also proves that some degrees of test anxiety can have positive effects on the writing performance of TOEFL iBT test takers. TOEFl iBT test anxiety, based on the results of the present study, is provoked by some factors like time limits, the length of the test and lack of self-confidence. TOEFL iBT test takers, therefore, need to learn and practice some time management strategies and test taking techniques to overcome test anxiety. The relations between kinesthetic intelligence and writing, and listening and musical intelligence also suggest that strategies like preparing an outline for a writing task in advance and paying attention to the parts and elements that convey important ideas, and the stress patterns sometimes used to put emphasis on them in a listening part can also help the candidates to improve their performance on the writing and listening sections of TOEFL iBT test.

However, the participants' performance on the speaking part of TOEFL iBT test was not investigated throughout the present study due to the complexities involved. Interested scholars can, therefore, do the same study with speaking in focus. They can also explore the effects of gender and examine the differences between male and female participants. It is also a good idea to do the same study with IELTS.

6.   References:

Armstrong, T. (2009). *Multiple intelligences in the classroom* (3rd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Assessment Reform Group. (2002). *Testing, motivation and learning* (Pamphlet No.ISBN 085603 046 5). Cambridge: University of Cambridge Faculty of Education.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. New York: Longman.

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

Brown, D. (2004). *Language assessment: Principles and classroom practice*. UK: Longman.

Brown, D. H. (1994). *Principles of language teaching and learning*. Englewood Cliffs: Prentice Hall Inc.

Cassady, C. J., & Johnson, E. R. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology, 27*, 270–295. http://dx.doi.org/10.1006/ceps.2001.1094

Chapell, M. S., Blanding, B. Z., Silverstein, M. E., Takahashi, M., Newman, B., Gubi, A., & McCann, N. (2005).Test anxiety and academic performance in undergraduate and graduate students. *Journal of Educational Psychology, 9*(2), 268-274. http://dx.doi.org/10.1037/0022-0663.97.2.268

Cumming, A., Kantor, R., Baba, K., Eouanzoui, K., Erdosy, U., & James, M. (2006). *Analysis of discourse features and verification of scoring levels for independent and integrated prototype written tasks for the new TOEFL*. Princeton, NJ: Educational Testing Service.

Fonseca-Mora, M. C., Toscano-Fuentes, C., & Wermke, K. (2011). Melodies that help: The relation between language aptitude and musical intelligence. *Anglistik International Journal of English Studies, 22*(1), 101-118.

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. USA: Routledge.

Gardner, R. (1985). *Social psychology and second language learning: The role of attitude and motivation*. London: Edward Arnold.

Gardner, R. (2001). Integrative motivation and second language acquisition. In Z. Dornyei, & R.W. Schmidt (Eds.), *Motivation and Second Language Acquisition* (pp. 1-19). University of Hawai'i: Honolulu.

Gregersen, T.S., & Horwitz, E.K. (2002). Language learning and perfectionism: Anxious and non-anxious language learners' reactions to their own oral performance. *The Modern Language Journal, 86*(3), 562-570. http://dx.doi.org/10.1111/1540-4781.00161

Horwitz, E. K., & Young, D. J. (Eds.). (1991). *Language anxiety: From theory and research to classroom implications*. Englewood Cliffs, NJ: Prentice-Hall.

Hsu, T.-C. (2012). A study on the EFL students' speech related anxiety in Taiwan. International *Journal of Research Studies in Language Learning, 1*(2), 3-18. http://dx.doi.org/10.5861/ijrsll.2012.v1i2.74

Kunnan, A. J. (2007). Test fairness, test bias, and DIF. *Language Assessment Quarterly, 4*(2), 109–112. http://dx.doi.org/10.1080/15434300701375865

Kunnan, A. J. (2004). Test fairness. In M. Milanovic, & C. Weir (Eds.), *European Year of Languages Conference Papers, Barcelona, Spain* (pp. 27- 48). Cambridge: Cambridge University Press.

Kunnan, A. J. (1995). Test taker characteristics and test performance. Cambridge: Cambridge University Press.

Lamb, M. (2004). Integrative motivation in a globalizing world. *System, 32,* 3-19. http://dx.doi.org/10.1016/j.system.2003.04.002

Loori, A. A. (2005). Multiple intelligences: A comparative study between the preferences of males and females. *Social Behaviour and Personality, 33*(1), 77-88. http://dx.doi.org/10.2224/sbp.2005.33.1.77

Lumley,T., & O'Sullivan, B. (2005). The effect of test-taker gender, audience and topic on task performance in tape-mediated assessment of speaking. *Language Testing, 22*(4), 415-437. http://dx.doi.org/10.1191/0265532205lt303oa

Mahdavy, B. (2008). The role of multiple intelligences (MI) in listening proficiency: A comparison of TOEFL and IELTS listening tests from an MI perspective. *Asian EFL Journal, 10*(3), 109-126.

Mahmoodzadeh, M. (2013). Investigating foreign language anxiety in Iranian classrooms: The effect of gender. *International Journal of Research Studies in Language Learning, 2*(1), 61-70. http://dx.doi.org/10.5861/ijrsll.2012.109

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (pp.13-103). New York: Macmillan.

Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13*, 242-256. http://dx.doi.org/10.1177/026553229601300302

Ohata, K. (2005). Potential sources of anxiety for Japanese learners of English: Preliminary case interviews with five Japanese college students in the U.S., *The Electronic Journal for English as a Second Language, 9*(3), 1 - 21.

Ohcubo, N. (2009). Validating the integrated writing task of the TOEFL internet based test (iBT): Linguistic analysis of test takers' use of input material. *Melbourne Papers in Language Testing, 14*(1).

Purpora, J. K. (2004). Validating questionnaires to examine personal factors in L2 test performance. In M. Milanovich, & C. Weir (Eds.), *European Language Testing in a Global Context. Proceedings of the Association of Language Testers of Europe (ALTE) Conference of Barcelona.* Cambridge: Cambridge University Press.

Razmjoo, S. A. (2008). On the relationship between multiple intelligences and language proficiency. The Reading Matrix, 8(2).

Richards, J. C., & Rodgers, T. S. (2001). *Approaches and methods in language teaching.* New York: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511667305

Rollinson, P. (2005). Using peer feedback in the ESL writing class. *English Language Teaching Journal, 59*(1), 23-30.

Saleem, H. (2008). Understanding kinesthetic intelligence. Retrieved January 22, 2014, from http://www.dirjournal.com/guides/understanding-kinesthetic-intelligence/

Sarason, I. G. (1980). *Test anxiety: Theory, research, and applications.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Sarason, I. G. (1986). Test anxiety, worry, and cognitive interference. In R. Schwarzer (Ed.), *Self-related cognition in anxiety and motivation* (pp. 19- 33). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sarason, I. G. (1984). Stress, anxiety, and cognitive interference: Reactions to tests. *Journal of Personality and Social Psychology, 46*, 929-938. http://dx.doi.org/10.1037/0022-3514.46.4.929

Skehan, P. (1989). *Individual differences in second language learning.* London: Edward Arnold.

Skehan, P. (1998). *A cognitive approach to language learning.* Oxford: Oxford University Press.

Takala, S., & Kaftandjieva, F. (2000). Test fairness: A DIF analyses of an L2 vocabulary test. *Language Testing, 17*, 323–340.

Tavakoli, M. (2009). The role of motivation in ESP reading comprehension test performance. *Journal of*

Teaching English Language and Literature Society of Iran (TELL), 9(3).

Tirri, K., & Nokelainen, P. (2011). *Measuring multiple intelligences and moral sensitivities in education*. Boston and Taipei: Sense Publishers Rotterdam. http://dx.doi.org/10.1007/978-94-6091-758-5

Visser, B. A., Ashton, M. C., & Vernon, P. A. (2006). Beyond g: Putting multiple intelligences theory to the test. *Intelligence, 34*, 487–502. http://dx.doi.org/10.1016/j.intell.2006.02.004

Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Basingstoke: Palgrave Macmillan.

Young, D. J. (1999). *Affect in foreign language and second language learning*. Boston, MA: McGraw-Hill.