



***Diagnosing L2 Learners' Language Skills Based on the Use of a  
Web-Based Assessment Tool Called DIALANG***

*Mahboubeh Taghizadeh, University of Tehran*

*Sayyed Mohammad Alavi, University of Tehran*

*Abbas Ali Rezaee, University of Tehran*

**Abstract**

This study was conducted with 68 Iranian students studying at the Alborz Institute of Higher Education. The participants' majors were Teaching English as a Foreign Language (TEFL;  $n = 23$ ), English Language Literature (ELL;  $n = 22$ ), and English Language Translation (ELT;  $n = 23$ ). DIALANG self-assessment scales, consisting of 107 statements with Yes/No responses, were used in this study. DIALANG is an online assessment system used for language learners who want to obtain diagnostic information about their language proficiency (Council of Europe, 2001). Results indicated that ELL students had the highest overall ranking, whereas TEFL students received the lowest overall ranking for listening skills. ELL students had the highest reading skill scores while ELT students demonstrated the lowest scores. ELL students ranked highest in writing ability whereas TEFL students rated lowest in writing skill. Kruskal-Wallis analysis revealed that there was no statistically significant difference in listening and reading skills across the three majors. One-way between-groups ANOVA did demonstrate a statistically significant difference in the writing self-assessment statements for the three groups. Implications and directions for future research with DIALANG are provided based on results from the study.

**Résumé**

Cette étude a été menée avec 68 étudiants iraniens qui étudient au Alborz Institute of Higher Education. Les matières principales des participants ont été l'enseignement de l'anglais comme langue étrangère (TEFL;  $n = 23$ ), la littérature d'expression anglaise (ELL;  $n = 22$ ), et la traduction de la langue anglaise (ELT;  $n = 23$ ). Les échelles d'auto-évaluation DIALANG, composées de 107 énoncés avec réponses oui/non, ont été utilisées dans cette étude. DIALANG est un système d'évaluation en ligne utilisé pour les apprenants qui veulent obtenir des informations diagnostiques de leur maîtrise de la langue (Conseil de l'Europe, 2001). Les résultats ont indiqué que les étudiants ELL avaient le rang global le plus élevé tandis que les étudiants TEFL ont reçu le plus bas classement général pour la capacité d'écoute. Les étudiants ELL avaient les scores les plus élevés des compétences en lecture tandis que les étudiants ELT ont démontré les scores les plus faibles. Les étudiants ELL ont obtenu le classement le plus élevé dans l'aptitude à écrire, tandis que les étudiants TEFL ont obtenu les notes les plus basses dans l'aptitude à écrire. L'analyse de Kruskal-Wallis a révélé qu'il n'y avait pas de différence statistiquement significative dans la capacité d'écoute et les compétences en lecture dans les trois matières principales. L'analyse simple de la variance entre les groupes ANOVA a démontré une différence statistiquement significative dans les énoncés d'auto-évaluation de l'écriture pour les trois groupes. Les répercussions et les orientations pour la recherche future avec DIALANG sont fournies en fonction des résultats de l'étude.



## Introduction

The introduction of alternative classroom assessment strategies in the early 1990s opened up new opportunities for language courses, language education, and language assessment (Esfandiari & Myford, 2013). Examples of alternative kinds of assessment include self-assessment (SA), peer-assessment, classroom observations by teachers, student portfolios, and interviews (Butler & Lee, 2010). Andrade, Du, and Mycek (2010) defined formative assessment as “a process during which students reflect on the quality of their work, judge the degree to which it reflects explicitly stated goals or criteria, and revise accordingly” (p. 3). Suzuki (2009) stated that the advantages of alternative assessment include the following: (a) quick administration; (b) students' involvement in the assessment process; (c) enhancement of students' autonomy in language learning; and (d) increase of students' motivation for language learning (Blanche & Merino, 1989; Brown & Hudson, 1998).

As technology evolves and online learning opportunities expand, language assessments such as those provided by DIALANG are easily administered. DIALANG is a Web-based assessment system used for language learners who want to obtain diagnostic information about their language proficiency (Council of Europe, 2001). DIALANG evaluates reading, writing, listening, grammar, and vocabulary skills. Speaking is not evaluated through DIALANG.

<http://www.lancaster.ac.uk/researchenterprise/dialang/about.htm>.

The purpose of this study was to examine language competency through the Web-based DIALANG language assessment tool across different majors at one university in Iran. The participants' majors were Teaching English as a Foreign Language (TEFL), English Language Literature (ELL), and English Language Translation (ELT). This study is important because there is a clear link between language proficiency and academic success (Sahragard, Baharloo, & Soozandehfar, 2011). While students were in the fourth year of their programs at the university where this study was undertaken, they had problems with comprehension of English language course materials and their language proficiency was less than adequate. This put these students at risk for not completing their programs and graduating from the university.

## LITERATURE REVIEW

### *Learner Self-Assessment*

Moritz (1996) considers self-assessment in foreign language education as a non-traditional form of assessment, and a logical component of both learner-centered pedagogies and more self-directed (autonomous) learning programs. Todd (2002) refers to self-assessment as an essential component in the learning experience of a self-directed learner. Conceptually, self-assessment is supported by theories of cognition, constructivism, and learner autonomy, especially those of Piaget and Vygotsky



(Chen, 2008). Deakin-Crick et al. (2005) also suggest that self-assessment builds on students' ownership of their language learning processes, self-awareness, and responsibility for the language learning experience. They further suggest that engaging learners in the assessment of their own language learning is related to theories of learning, acceptance of the importance of motivation for learning, and the value of non-cognitive results. Similarly, Boud (1995) notes that self-assessment originated in the context of autonomous learning or learner independence. Brown (2004) also suggests that the principles of autonomy, intrinsic motivation, and cooperative learning comprise the theoretical justifications for self-assessment. Similarly, Alderson and McIntyre (2006) argue that implementation of self-assessment arises out of belief in student autonomy as an educational goal.

A great number of advantages have been identified for self-assessment including the following: raising the level of awareness about the learning process (Benson, 2001; Blanche & Merino, 1989; Kato, 2009; Oscarson, 1989; Todd, 2002); promotion of learner autonomy (Cram, 1995; Dann, 2002; Kato, 2009; Oscarson, 1989, 1997; Paris & Paris, 2001); setting of realistic goals and directing personal learning (Abolfazli Khonbi & Sadeghi, 2012; Blanche & Merino, 1989; Butler & Lee, 2010; Oscarson, 1989); discernment of individual patterns of strengths and weaknesses (Blue, 1994; Esfandiari & Myford, 2013; Saito & Fujita, 2004); increasing learner motivation (Barbera, 2009; Paris & Paris, 2001; Sadler & Good, 2006; Todd, 2002); increased effect on students' learning over time (Oscarson, 1989). Other benefits include expansion of assessment types (Oscarson, 1989); monitoring personal progress and reflection on what should be done (Barbera, 2009; Butler & Lee, 2010; Esfandiari & Myford, 2013; Hana Lim, 2007; Haris, 1997; Peden & Carroll, 2008; Sadler & Good, 2006; Sally, 2005); facilitation of democratic learning processes (Oscarson, 1989; Shohamy, 2001); taking responsibility for learning (Barbera, 2009; Esfandiari & Myford, 2013; Paris & Paris, 2001; Peden & Carroll, 2008; Sadler & Good, 2006); and promotion of learning (Black & Wiliam, 1998; Oscarson, 1989).

In addition, the factors reported to influence the implementation of self-assessment are clear criteria (Airasian, 1997; Falchikov, 1986; Orsmond et. al, 2000; Stiggins, 2001); training before the actual assessment (AlFallay, 2004; Chen, 2006; Wiggins, 1993); sufficient practice (McDonald & Boud, 2003; Nicol & Macfarlane-Dick, 2006; Orsmond et al., 2000; Stefani, 1998; Taras, 2001). Teacher intervention and feedback (Orsmond et al., 2002; Stanley, 1992; Taras, 2003) and cultural and educational context (Oscarson, 1997) are also factors reported in the literature.

Of the techniques for involving learners in self-assessment, objectively-marked discrete-point tests of linguistic knowledge, rating scales, and checklists are three traditional approaches to self-assessment of language ability (Brindley, 1989; North, 2000; Oscarson, 1989). Blanche (1988) also identifies techniques such as checklists (e.g., questionnaires /'can-do' statements), learners' diaries, learners' reports on real-life communication, self-ratings of certain instructional objectives, and retrospective self-assessment where learners report on their success or lack of success when communicating with native speakers



outside the classroom and in other contexts. Of all these techniques, rating scales with holistic descriptors are the most commonly used self-assessment technique (Brindley, 1989; North, 2000).

The relationship between self-assessment and learning and teaching contexts is another issue. Since self-assessment can potentially modify the power relationship between students and teachers, some teachers may find it a challenge to their authority (Towler & Broadfoot, 1992). Hamp-Lyons (2007) described two conflicting cultures of assessment: an exam culture and a learning culture. A focus in a learning culture is individual learners' improvement in learning, while an exam culture concentrates on learners' mastery of language proficiency in relation to norms or groups. Hamp-Lyons further states the transition from an exam culture to a learning culture is a complex process, requiring consideration of teachers' viewpoints in order to make the transition successful.

Closely related to the problem of overestimation, as Esfandiari and Myford (2013) argue, is that of the accuracy, validity and reliability of self-ratings may not be an accurate reflection of individual abilities. Chen (2008) states that "literature on student self-assessment often discusses its validity and reliability, or effectiveness, as an assessment tool based on agreement between self- and teacher scorings" (p. 253). For instance, Topping (2003) notes that (a) scores that students gave to themselves seemed to be higher than scores that teachers gave to them, and (b) self-assessment based on the students' perceptions of their levels of effort rather than their levels of achievement were particularly unreliable. Additionally, self-assessments appear to be more unreliable when students rate their own performance than when they assess their own learning products (Segers & Dochy, 2001). However, Bachman and Palmer (1989) argue that self-assessment tends to be a reliable and valid instrument for communicative language ability though learners at a lower level compared to the more proficient ones may find the self-assessment difficult. Likewise, LeBlanc and Painchaud (1985) and Pierce et al. (1993) note that self-assessment tends to be a valuable and reliable indicator of language proficiency.

According to Butler and Lee (2010), the inherent subjectivity of self-assessment as a measurement tool has traditionally been reported as a threat to its validity. As a result, research analyzing the measurement aspect of self-assessment in foreign and second language education has focused on the validity of self-assessment (Butler & Lee, 2010). Butler and Lee further state that, "such validation studies have often examined the correlations between self-assessment scores and scores obtained through various types of external measurements such as objective tests, final grades, and teachers' ratings" (p. 7). In studies on validation (Blanche & Merino, 1989; Oscarson, 1997; Ross, 1998), the results are mixed, and a number of factors have been identified regarding the variability in self-assessment results which can be broadly categorized as follows: "(1) the domain or skill being assessed; (2) students' individual characteristics; and (3) the ways in which questions and items are formulated and delivered" (p. 7).



### **Web-Based Language Testing**

With the expansion of distance and online learning, rapid development of IT and the Internet, and increasing accessibility and availability of both hardware and software, computers are now widely used in language education (Davies, 2003) and are an increasingly important factor of change in education (Alvarez & Rice, 2006). For instance, as Micea (2005) points out, e-learning can be viewed as a means of facilitating three significant outcomes: increased equity, improved productivity and improved innovation and competitiveness, and improved and consistent rates of lifelong learning. In this light, Warschauer and Meskill (2000) argue that, by using new technologies in the language classroom, teachers can better prepare students for the kinds of international cross-cultural interactions that are increasingly required for success in academic, vocational, and personal life.

The use of computer technology in teaching languages has dramatically increased worldwide over the past decade (Chen, Belkada, & Okamoto, 2004; Hubbard & Levy, 2006; Son, 2008). As a result, a fairly large literature exists on the effectiveness of computer-assisted language learning on language development. The findings of these studies, according to Silye and Wiwczarowski (2002), suggest that both language learners and instructors have generally positive attitudes toward using computers in the language classroom. Less is known, however, about the more specific use of computers in the language testing area.

The Web, as a recently emerged instrument of language assessment (Silye, & Wiwczarowski, 2002), greatly expands the availability and accessibility of computer-based testing with all its potential advantages. Additionally, it will undoubtedly become a main medium of test delivery in the near future. Silye and Wiwczarowski further argue that a Web-based test is an assessment instrument that is written in the "language" of the Web, HTML. The test itself consists of one or several HTML file(s) located on the tester's computer and the server. It can be downloaded to the test taker's or client's computer. Downloading can occur for the entire test at once, or item by item. The client's computer makes use of Web-browser software such as Google Chrome or Microsoft Internet Explorer, to interpret and display the downloaded HTML data. Test takers respond to items on their computers and may send their responses back to the server. Alternately, their responses can be scored in real time by means of a scoring script administered by the person providing oversight for the test. A script can then be generated to provide immediate feedback, adaptation of item selection to the test taker's needs, and/or computation of a score to be displayed after completion of the test. The same evaluation process can take place on the server by means of server-side programs.

It is obvious that there are various advantages linked to the use of Web-based tests. Roever (2001) suggests three main advantages: (a) flexibility in space and time, which is probably the biggest advantage of a Web-based test, since all that is required to take a Web-based test is a computer with a Web browser and an Internet connection. In addition, test takers can take this test wherever and



whenever it is convenient, and test developers can share their test with colleagues all over the world and receive feedback. (b) Web-based tests are comparatively easy to write and require only a free, standard browser for their display; and (c) a Web-based test is very inexpensive for all parties concerned, including testers and test takers. Alvarez and Rice (2006) argue that Web-based tests provide immediate feedback, improved security, ways to store test results for further analysis, storage of large amounts of items, multimedia presentations, grading objectivity, and self-pacing for the test taker.

Despite the numerous advantages of Web-based tests, some limitations are suggested. For instance, Silye and Wiwczarowski (2002) contend that the greatest limitation of these types of tests is their lack of security with regard to item confidentiality and cheating. Alvarez and Rice (2006) suggest security and/or technical problems such as browser incompatibilities and server failure as other drawbacks in the use of Web-based tests. They further note that computers lack human intelligence to assess direct speaking ability and freely written compositions.

#### ***DIALANG as a Web-Based Assessment Tool***

DIALANG, as Alderson (2005) states, is a computer-based learner-centered diagnostic language assessment system, which is intended for language learners who want to obtain diagnostic information about their proficiency. Like other Web-based tests, DIALANG offers many advantages such as flexibility in space and time, easiness to write, and affordability (Alvarez & Rice, 2006). Brantmeier and Vanderplank (2008) assert that for the DIALANG project, Alderson specifically excludes self-assessment from high stakes testing and regards self-assessment more as a valuable descriptive and explanatory tool for providing feedback to learners.

DIALANG is a large and complex project that is funded by the European Union (Escribano & McMahon, 2010) and developed by a team of experts at the University of Lancaster (Klimova & Hubackova, 2013). According to Chapelle (2006), theoretical and empirical rationales have been taken into account in the design and development of the DIALANG. DIALANG, as Alderson (2005) states, contains tests of five language skills or aspects of language knowledge (i.e., Listening, Reading, Writing, Grammar, and Vocabulary) and due to the constraints on the computer-based testing, the CEFR scales for spoken production were ignored (Alderson, 2005).

DIALANG, according to Alderson (2005), is unique in that it attempts the diagnostic assessment of 14 European languages: Danish, Dutch, English, Finnish, French, German, Greek, Italian, Portuguese, Spanish, Swedish, Irish, Icelandic, and Norwegian. DIALANG's Assessment Framework and the descriptive scales used for reporting the results to the users are directly based on the CEFR (Council of Europe, 2001). DIALANG Framework, as Haahr and Hansen (2006) assert, summarizes "the relevant content of the CEFR, including the six-point reference scale, communicative tasks and purposes, themes



and specific notions, activities, texts and functions" (p. 78). They further state that DIALANG provides learners with various kinds of feedback on the weak and strong points in their language proficiency and constructive advice for further learning. Diagnosis offered in DIALANG is not concerned with specific language curricula or courses; rather, it is based on the specifications of language proficiency suggested in the CEFR.

According to Haahr and Hansen (2006), the main methodological steps and challenges in DIALANG's development of a standard based on the CEFR are (1) awareness of the theoretical assumptions of the framework; (2) complementing for the limitations of the framework; (3) explaining the framework to the assessment development teams; (4) conservative or innovative items? (i.e., do the test items represent a variety of different formats? or are the possibilities of computer-based tests sufficiently utilized?); (5) piloting of items; (6) relating test items and resulting scores to the CEFR levels (p. 78).

DIALANG serves adults who want to know about their levels of language proficiency and who want to receive feedback on the weaknesses and strengths of their proficiency (Council of Europe, 2001). According to Alvarez and Rice (2006), the main advantage of the DIALANG is perhaps the way in which this project provides assessment information: (a) scores are objective, (b) the feedback is immediate, (c) instead of a global score, there is identification of the test taker's current language level, (d) it enables the planning of curricula, (e) it offers learners' further study opportunities, (f) it meets the learner's need for feedback, and (g) it allows for storage of outcomes for later comparisons in order to check progress. Alderson (2005) suggests that the value of the DIALANG Framework is in the following areas:

- It has been possible to gather information from a fairly large number of candidates, at least for English, from a wide range of national and linguistic as well as educational backgrounds. (p. 29)
- The DIALANG system aims to encourage self-diagnosis and self-help and is entirely learner-centered (p. 31)
- For the learner or student, the DIALANG system is a free of charge, no stakes test. (p. 31).

By contrast, in Brunfaut (2014), Alderson suggests two areas in which the DIALANG has not been successful: (a) in some of the languages involved in the program, items are limited, and (b) the theory behind this test is fairly traditional. Alvarez and Rice also argue that a real challenge for Web-based tests is assessment of productive skills (i.e., speaking and writing). With respect to DIALANG, they state that, at the moment, DIALANG lacks a system to assess writing and listening in terms of full sentences, paragraphs, and essays. They further note that a project to consider these deficiencies is currently taking place, and they present some examples of the items produced in the experimental phase of this project.



The system can be downloaded from the DIALANG Website at [www.dialang.org](http://www.dialang.org). Users have to be connected to the Internet to take the tests because test items and interface texts come from the DIALANG servers and are cached on the individual user's machine (Alderson & Huhta, 2005). Alderson and Huhta (2005) and Alvarez and Rice (2006) describe the procedure for taking the DIALANG test as follows: the procedure to follow in order to take the test does not need a lot of computer skills. The first thing to do is to choose the language in which instructions will be given. After that, the language and skill to be tested are selected. Then, test takers have the option to take a vocabulary placement test. The system uses the vocabulary test score to select a test of suitable difficulty for the test taker. The next step is an optional battery of self-assessment statements about listening, reading, and writing skills before proceeding to the test itself. If users have completed both the vocabulary placement test and the self-assessment, the system combines the two results to decide which level of test to administer. Next, learners take the test for the skill chosen previously. The items are frequently multiple choice questions but sometimes a word or phrase has to be written. Finally, DIALANG presents the results in terms of the CEFR level, answer verification, score in the placement test, self-assessment feedback, and advice.

Self-assessment, as Alderson (2005) notes, is the central component of DIALANG, and the self-assessment statements are taken directly from the CEFR; specifically, they include a large number of 'can do' statements for each skill and at each level (Haahr & Hansen, 2006). However, the wording of CEFR statements was changed from 'can do' to 'I can' while some statements were also simplified for use with certain audiences. DIALANG also developed a number of 'can do' statements for grammar and vocabulary (Haahr & Hansen, 2006).

As Haahr and Hansen (2006) point out, the self-assessment statements in the DIALANG Framework underwent a piloting procedure similar to the test items and the correlations between their calibrated values of difficulty and the original CEFR levels were very high (0.91-0.93). This result indicates that the DIALANG self-assessment statements correspond closely to the original CEFR levels. It was also revealed that the self-assessment statements were equivalent across different languages (Alderson, 2005).

### ***Research Aims, Purpose and Questions***

While research has been conducted on self-assessment in language learning (Alderson, & McIntyre, 2006; Chen, 2008; Deakin-Crick et al., 2005; Escibano & McMahan, 2010; Hana Lim, 2007; Kato, 2009; Suzuki, 2009; Wagner & Lilly, 1999), very little empirical examination has been undertaken using DIALANG to assess language proficiency, particularly in the Iranian context. The purpose of this study was, therefore, to examine second language (L2) learners' self-assessed level in listening, reading, and writing skills in terms of SA statements of the DIALANG project:

1. Is there any statistically significant variability in the L2 learners' performance in the listening section of the DIALANG scale?





2. Is there any statistically significant variability in the L2 learners' performance in the reading section of the DIALANG scale?
3. Is there any statistically significant variability in the L2 learners' performance in the writing section of the DIALANG scale?

## METHODS

### **Participants**

A convenience sample of 68 Iranian Bachelor of Arts students studying at an institute of higher learning in Iran participated in the study. The majors of the participants were Teaching English as a Foreign Language (TEFL;  $n = 23$ ), English Language Literature (ELL;  $n = 22$ ), and English Language Translation (ELT;  $n = 23$ ). Participants were male (27%) and female (73%) and ranged in age from 18 to 27. Participants were all in their fourth year of study. Approval to conduct the study was obtained at the university prior to data collection.

### **Instruments**

DIALANG scales of self-assessment were used in this study. They consisted of 107 statements with “yes” or “no” responses. The underlying constructs of the scales included reading skill (31 self-assessment statements), listening skill (43 self-assessment statements), and writing skill (33 self-assessment statements). Based on the Common European Framework of Reference for Languages (CEFR) (<https://www.eui.eu/Documents/ServicesAdmin/LanguageCentre/CEF.pdf>), the levels are understood as follows: at the A level, the learner is considered to be at the basic level of proficiency; at the B level, the learner is considered to be at the independent user level; at the C level, the learner is considered to be at a proficient level. More information regarding the CEFR can be found in Appendix A. Detailed information about the different levels of each DIALANG construct is presented in Table 1.

Cronbach's alpha was used to estimate the consistency of the participants' responses to the DIALANG scales. Reliability coefficients for the listening, reading, and the writing scales were 0.90, 0.88, 0.88, respectively, indicating that the scales demonstrated high reliability.

Table 1. The Number of Self-Assessment Statements for the Levels of the DIALANG Scales

Skills	Levels						Total
	A1	A2	B1	B2	C1	C2	
Listening	4	10	10	9	9	1	43
Reading	5	9	8	6	2	1	31
Writing	6	7	7	4	5	4	33

### Procedures

The study was conducted at the end of the Fall Semester in 2013. The researchers provided the participants with some background about the DIALANG self-assessment experience and explained the goal of the study in order to encourage their participation. Then, the Persian version of the DIALANG self-assessment scales was administered to the students. They were asked to complete the self-assessment in 100 minutes.

### Statistical Analysis

To answer the research questions, various statistical analyses were performed. Descriptive statistics and chi-square analysis for all self-assessment statements of the DIALANG scales were performed. In addition, descriptive statistics for listening, reading, and writing skills and participants' levels were calculated for the TEFL, ELL and ELT students. Further, to investigate for a statistically significant difference in the participants' responses to the three parts of DIALANG scales, Kruskal-Wallis Test and ANOVA were performed.

## RESULTS

### Statements on DIALANG Scales

In order to determine which items received more positive replies, the frequency and percentage of the participants' responses for each item in the listening, reading, and writing skills scales were calculated. With regard to listening skill, Item 3, 'Understanding questions and instructions and following short, simple directions' generated the highest frequency ( $f = 65$ ), while Item 42, 'Following films which contain a considerable degree of slang and idiomatic usage,' generated the lowest frequency ( $f = 32$ ).



Considering the chi-square and p-values of the listening scale statements, the frequency occurrences of the participants' yes/no responses were significantly different in most of the items except the following: Item 30, 'Understanding announcements and messages on concrete and abstract topics spoken in standard language at normal speed' ( $\chi^2 = 2.12$ ,  $p = .146$ ); Item 37, 'Following extended speech even it is not clearly structures and when relationships between ideas are only implied and not stated explicitly' ( $\chi^2 = 0.00$ ,  $p = 1.000$ ); Item 38, 'Following most lectures, discussions and debates with relative ease' ( $\chi^2 = 0.00$ ,  $p = 1.000$ ); Item 39, 'Extracting specific information from poor quality public announcements' ( $\chi^2 = 2.12$ ,  $p = .146$ ); Item 41, 'Understanding a wide range of recorded audio material and identifying finer points of detail' ( $\chi^2 = 2.12$ ,  $p = .146$ ); Item 42, 'Following films which contain a considerable degree of slang and idiomatic usage' ( $\chi^2 = 0.24$ ,  $p = .628$ ); and Item 43, 'Following specialized lectures and presentations which use a high degree of colloquialism, regional usage or unfamiliar terminology' ( $\chi^2 = 0.06$ ,  $p = .808$ ). The frequency occurrences of participants' yes/no responses were not statistically different. This finding indicates that, statistically, the participants held the same viewpoints regarding these statements.

Related to reading skill, the two statements of A1 level: Item 2, 'Understanding very short, simple texts, putting together familiar names, words, and basic phrases', and Item 3, 'Following short, simple written instructions, especially if they contain pictures', and two statements of A2 level: Item 7, 'Understanding short, simple texts written in common everyday language' and Item 11, 'Understanding short, simple personal letters' received the highest frequency ( $f = 64$ ), whereas the only statement of C2 level, 'Understanding and interpreting practically all forms of written language', received the lowest frequency ( $f = 34$ ). In addition, the frequency occurrences of participants' yes/no responses were significantly different in 28 statements, except for Item 27, 'Identifying the content and relevance of news items, articles and reports on a wide range of professional topics' ( $\chi^2 = 3.76$ ,  $p = .052$ ), Item 30, 'Understanding long, complex instructions on a new machine or procedure even outside area of specialty' ( $\chi^2 = 3.76$ ,  $p = .052$ ), and Item 31, 'Understanding and interpreting practically all forms of written language' ( $\chi^2 = 0.00$ ,  $p = 1.000$ ).

With regard to writing statements, 'Writing simple notes to friends' and 'Writing very simple personal letters expressing thanks and apology' experienced the highest frequency ( $f = 65$ ). Item 31, 'Producing clear, smoothly flowing, complex reports, articles or essays' received the lowest frequency. The frequency occurrences of the participants' yes/no responses were significantly different in most of the items except for Item 23, 'Constructing a chain of reasoned argument' ( $\chi^2 = 0.53$ ,  $p = 0.467$ ); Item 24, 'Speculating about causes, consequences and hypothetical situations' ( $\chi^2 = 0.94$ ,  $p = .332$ ); Item 25, 'Expressing and supporting points of view at some length with subsidiary points, reasons and relevant examples' ( $\chi^2 = 0.53$ ,  $p = .467$ ); Item 26, 'Developing an argument systematically, giving appropriate emphasis to significant points, and presenting relevant supporting details' ( $\chi^2 = 0.53$ ,  $p = .467$ ); Item 27, 'Giving clear detailed descriptions of complex subjects' ( $\chi^2 = 0.00$ ,  $p = 1.000$ ); Item 30, 'Providing an



appropriate and effective logical structure' ( $\chi^2 = 2.12, p = .146$ ); Item 31, 'Producing clear, smoothly flowing, complex reports, articles or essays' ( $\chi^2 = 0.94, p = .332$ ); Item 32, 'Writing so well that native speakers need not check my texts' ( $\chi^2 = 0.53, p = .467$ ); and Item 33, 'Writing so well that my texts cannot be improved significantly even by teachers of writing' ( $\chi^2 = 0.53, p = .467$ ).

**Comparison of Scales Across the TEFL, ELL, and ELT Majors**

Descriptive statistics for the listening scale levels for the three majors are reported in Table 2.

Table 2. Means (M) and Standard Deviations (SD) for Listening Scale Levels Based on Major

Levels	Major					
	TEFL		ELL		ELT	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
A1	3.65	0.77	3.86	0.35	3.83	0.38
A2	9.17	1.72	9.09	1.19	8.78	2.06
B1	8.87	1.74	9.18	1.81	8.09	2.71
B2	6.13	2.13	6.73	2.88	5.61	3.05
C1	4.13	2.89	6.91	2.74	5.48	2.84
C2	0.35	0.48	0.50	0.51	0.70	0.47

Students from different majors did not report equal capabilities about the 'can do' statements of the listening scale. The highest capabilities ( $M = 9.18$ ) were reported by the students of ELL for the B1 statements, while the lowest ( $M = 0.35$ ) were reported by the TEFL students for the statements of the C2 level. The ELL students' responses to the statements of the A1 level were the most homogeneous ( $SD = 0.35$ ), while the ELT students' responses to the B2 level statements were the most heterogeneous ( $SD = 3.05$ ).

Further, ELL students reported that they were more at ease with 'can do' statements of the A1 level, while TEFL students reported lowest abilities ( $M = 3.65$ ). With regard to A2 Level statements, TEFL students gained the highest mean score ( $M = 9.17$ ), whereas ELT students received the lowest mean



score ( $M = 8.78$ ). Related to B1 statements, the mean score for the ELL students was the highest ( $M = 9.18$ ), whereas the mean score for the ELT students was the lowest ( $M = 8.09$ ). Similarly, ELL students reported more capabilities ( $M = 6.73$ ) in the statements at the B2 level, while ELT students reported lowest abilities ( $M = 5.61$ ). The highest proportion of the ELL students' overall positive replies ( $M = 6.91$ ) was reported for the statements of C1 level; the lowest one ( $M = 4.13$ ) was reported by the TEFL students. Concerning statements of the C2 level, ELT students received the highest mean score ( $M = 0.70$ ), while TEFL students received the lowest mean score ( $M = 0.35$ ). Table 3 shows that learners had different abilities with respect to the six levels for the reading scale.

Table 3. Means (M) and Standard Deviations (SD) for Reading Scale Levels Based on Major

Levels	Major					
	TEFL		Literature		Translation	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
A1	4.61	1.07	4.73	0.63	4.52	0.94
A2	7.87	1.93	8.00	1.71	7.09	2.59
B1	6.96	1.69	7.09	2.97	5.87	2.43
B2	4.04	1.63	4.36	2.03	3.83	2.14
C1	1.35	0.77	1.50	0.67	1.43	0.72
C2	0.39	0.49	0.55	0.51	0.57	0.50

ELL students gained the highest mean score ( $M = 8.00$ ) in the statements of A2 level, while TEFL students received the lowest mean score ( $M = 0.39$ ) in the C2 statements. The most homogeneous responses ( $SD = 0.49$ ) occurred with statements of the C2 level as reported by TEFL students, whereas the most heterogeneous ones ( $SD = 2.97$ ) were related to the ELL students.

At the A1 level, the mean score for the ELL students was highest ( $M = 4.73$ ), while scores for the ELT students was lowest ( $M = 4.52$ ). The same trend appeared for the statements of the A2, B1, and the C1 levels. In other words, ELL students reported highest abilities in these levels, while ELT students reported lowest abilities: A2:  $M_L = 8.00$ ,  $M_T = 7.09$ ; B1:  $M_L = 7.09$ ,  $M_T = 5.87$ ; B2:  $M_L = 4.36$ ,  $M_T =$



3.83. However, at the C1 level, ELL students reported to be most capable ( $M = 1.50$ ), while TEFL students reported lowest abilities. For statements at the C2 level, ELT students gained the highest mean score ( $M = 0.57$ ), while TEFL students received the lowest one ( $M = 0.39$ ). Table 4 shows that learners had different abilities with respect to the six levels for the writing scale.

Table 4. Means (M) and Standard Deviations (SD) for Writing Scale Levels Based on Major

Levels	Major					
	TEFL		Literature		Translation	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
A1	5.52	0.94	5.55	0.96	5.74	2.66
A2	6.39	1.03	5.91	1.68	5.48	1.88
B1	5.83	1.37	5.82	1.86	4.91	2.39
B2	2.17	1.07	2.77	1.34	2.43	1.75
C1	2.39	1.40	3.14	1.80	3.22	1.97
C2	1.30	1.36	2.36	1.64	2.17	1.64

TEFL students rated their writing abilities at the A2 level the highest ( $M = 6.39$ ), while their self-ratings of the C2 level statements received the lowest mean score ( $M = 1.30$ ). The most homogeneous responses ( $SD = 0.94$ ) were related to the A1 level among TEFL students' responses, while the most heterogeneous ones were reported by ELT students at the A1 level. Overall, the participants of this study assessed their writing ability in the following order: TEFL (A2, B1, A1, C1, B2, C2); ELL students (A2, B1, A1, C1, C2, B2); ELT students (A1, A2, B1, C1, B2, C2).

#### **Participants' Total Score on the Three Scales of DIALANG**

Before investigating whether there were any statistically significant differences in the listening and reading skills scales for the three groups, tests of normality were conducted. Results showed that the listening and reading scores violated the assumption of normality ( $p < .05$ ). Therefore, in order to compare the participants relative to these two scales, a Kruskal-Wallis (K-W) test was conducted. Results are presented in Tables 5 and 6.

Table 5. K-W Test for the Listening Scale Statements

Skill	Major	N	Mean Rank	Chi-Square	df	p
	TEFL	23	28.63	5.366	2	.068
Listening	Literature	22	42.05			
	Translation	23	33.15			

ELL students had the highest overall ranking (Mean rank = 42.05), whereas TEFL students received the lowest overall ranking (Mean Rank = 28.63). In addition, a K-W test did not reveal a statistically significant difference in the listening statements across the three majors (TEFL group, n = 23, ELL group, n = 22, ELT group, n = 23),  $\chi^2(2, n = 68) = 5.366, p = .068$ .

Table 6. K-W Test for the Reading Scale Statements

Skill	Major	N	Mean Rank	Chi-Square	df	p
	TEFL	23	34.48	1.187	2	.552
Reading	Literature	22	37.77			
	Translation	23	31.39			

Results indicated that ELL students had the highest scores (Mean Rank = 37.77), with the ELT students reporting the lowest scores (Mean Rank = 31.39). The results of K-W test showed that there was no statistically significant difference in the reading statements across the three majors (TEFL group, n = 23, ELL group, n = 22, ELT group, n = 23),  $\chi^2(2, n = 68) = 1.187, p = .552$ .

The normality tests for the writing scale were also assessed. Results indicated that the writing scores of all majors did not violate the assumption of normality ( $p > .05$ ). Therefore, in order to compare the participants' viewpoint with regard to writing statements, ANOVA test was conducted. ANOVA results are presented in Table 7.

Table 7. Means and Standard Deviations of Students' Self-Assessment of their Writing Ability

	TEFL		Literature		Translation	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Writing	23.61	3.76	25.55	6.80	23.96	9.51

ELL students reported more capabilities based on the writing statements, whereas TEFL students reported the less abilities ( $M = 23.61$ ) in this regard. Table 7 also shows that the most homogeneous responses were generated by the TEFL students, while the most heterogeneous ones were from ELT students. As described in Table 8, one-way between-groups ANOVA did not show a statistically significant difference in the writing self-assessment statements for the three groups.

Table 8. Comparing the Participants' Scores on the Writing Scale

Writing	<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>p</i>
Between Groups	47.640	2	23.820	.473	.625
Within Groups	3275.889	65	50.398		
Total	3323.529	67			

### DISCUSSION

In spite of the six-year compulsory English education at the secondary level, and at least three years of English language courses at the post-secondary level, the undergraduate students in this study did not report high language skill ability, and their scores for listening, reading and writing were quite low. The researchers believe that these results may be related to a few factors. First, before entering university, language courses are generally focused on reading, and no systematic instruction is offered to master listening, writing, or speaking skills. Large class sizes may be another reason for the low scores, because students may not receive the individualized instruction and attention they need to develop language skills.





In addition, traditional teacher-centered teaching methods, lack of standards for language proficiency, and the lack of predetermined, concrete learning outcomes may account for the problems with English language instruction at this university. Moreover, students' assessment is mostly summative, in which learning outcomes are reported as a single score and as a result, they are not often asked to evaluate their abilities or to be involved in the self-assessment in their language learning process.

Technology and online learning opportunities may provide additional learning opportunities for students at this university, however, both are currently under-utilized. In the absence of technology and Web-based learning opportunities, an alternative means for assessing language proficiency and early identification of deficiencies that is provided by DIALANG may be warranted. The researchers believe that DIALANG self-assessment may help students in this university take control of their language learning because of its many features. Features offered by DIALANG include a learner feedback and identification of language deficiencies, which can help learners diagnose their weaknesses and strengths and increase their awareness of their language levels.

### **Study Limitations**

The small sample size makes findings not generalizable beyond this post-secondary institution where the study was conducted. Further, given the researchers' relationships with participants as colleagues and administrators, it is not known whether responses were guarded or offered in order to be socially desirable. Participant availability was another limitation of this study. Additionally, factors such as sociocultural background, L2 proficiency level, gender, and age were not taken into account and may have provided additional insights about why particular scores were obtained.

### **Conclusion**

This study aimed to investigate L2 learners' level in the listening, speaking, and writing skills in terms of 'can do' statements of the DIALANG Framework. To this end, undergraduate students of the English language were asked to rate their abilities based on the self-assessment statements of DIALANG. The results of the study revealed that the participants of the study rated their language abilities in the following order: TEFL students (reading, listening, writing); ELL students (listening, reading, writing); and ELT students (listening, reading, writing). Although ELL students gained the highest mean score for all three skills, there was not any statistically significant difference in the reading, writing, and listening statements across the three majors.

Language teachers in both online and face-to-face classes should promote learners' ability to self-assess and to reflect on their language proficiency level in order to help them facilitate the process of language learning and formulate and analyze the steps needed to achieve the learning goals. It is suggested that teachers implement self-assessment and introduce DIALANG statements as part of language instruction, and they train students to conduct self-assessment based on the 'can do' statements. Students in both language learning environments can also evaluate their progress in the language skills based on the 'can do' statements and then can formulate specific goals for their future progress. They can examine how well they have attained the learning goals for the course based on the 'can do' statements. To improve



learners' language skills, materials developers can also consider language proficiency standards offered in the DIALANG Framework to develop teaching materials and textbooks, as it is believed that DIALANG provides learners with concrete learning outcomes, which is one of the essential issues in the course design.

It is suggested that research be conducted comparing the language proficiency level of both online students and those who study in traditional language classes. After that, based on the results of the DIALANG test, a comparison can be made in terms of students' strengths and weaknesses in these two educational environments. In future study, learners in both face-to-face and online language classes can receive training based on the 'can do' statements and then put in charge of rating their own performance. Then, the impact of this training on the language proficiency of these learners can be compared. The relationship between self-assessment in terms of DIALANG statements and factors such as personality traits, learning anxiety, locus of control, and the cognitive style merits further inquiry. Future researchers can also use more qualitative and in-depth interview with learners and tutors about L2 performance with respect to DIALANG test and its 'can do' statements.



**APPENDIX A**

Proficient User	C2	Can understand with ease virtually everything heard or read. Can summarize information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
	C1	Can understand a wide range of demanding, longer texts, and recognize implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organizational patterns, connectors and cohesive devices.
Independent User	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialization. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.
Basic User	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

Obtained on 11-18-14 from: <https://www.eui.eu/Documents/ServicesAdmin/LanguageCentre/CEF.pdf>



#### REFERENCES

- Abolfazli Khonbi, Z., & Sadeghi, K. (2012). The effect of assessment type (self vs. peer) on Iranian university EFL students' course achievement. *Procedia - Social and Behavioral Sciences*, 70, 1552 – 1564.
- Airasian, P. W. (1997). *Classroom assessment* (3rd ed.). New York: McGraw-Hill.
- Alderson, Ch. J. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.
- Alderson, Ch. J., & McIntyre, D. (2006). Implementing and evaluating a self-assessment mechanism for the Web-based language and style course. *Language and Literature*, 15(3), 291–306.
- AlFallay, I. (2004). The role of some selected psychological and personality traits of the rater in the accuracy of self- and peer- assessment. *System*, 32(3), 407–425.
- Alvarez, M., & Rice, J. (2006). Web-based tests in second/foreign language self-assessment. The 29th Annual Proceedings, 2, (pp.13-21). Dallas: *The National Convention of the Association for Educational Communications and Technology*.
- Andrade, H., Du, Y., & Mycek, K. (2010). Rubric-referenced self-assessment and middle school students' writing. *Assessment in Education*, 17(2), 199-214.
- Bachman, L., & Palmer, A. (1989). The construct validation of self-ratings of communicative language ability. *Language Testing*, 6, 14–29.
- Barbera, E. (2009). Mutual feedback in e-portfolio assessment: An approach to the net folio system. *British Journal of Educational Technology*, 40(2), 342-357.



- Benson, P. (2001). *Teaching and researching autonomy in language learning*. London: Longman.
- Black, P., & Wiliam, D. (1998.). Assessment and classroom learning. *Assessment in Education*, 5, 7-74.
- Blanche, P. (1988). Self-assessment of foreign language skills: Implications for teachers and researchers. *RELC Journal*, 19(1), 75–96.
- Blanche, P., & Merino, B. J. (1989). Self-assessment of foreign language skills. *Language Learning*, 39(3), 313–340.
- Blue, G. (1994). Self-assessment of foreign language skills: Does it work? *CLE Working Papers*, 3, 18–35.
- Boud, D. (1995). *Enhancing learning through self-assessment*. London: Kogan Page.
- Brantmeier, C., & Vanderplank, R. (2008). Descriptive and criterion-referenced self-assessment with L2 readers. *System*, 36, 456–477.
- Brindley, G. (1989). *Assessing achievement in the learner-centered curriculum*. NCELTR: Sydney, NSW.
- Brown, H. (2004). *Language assessment: Principles and classroom practices*. New York, NY: Longman.
- Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32, 653-675.
- Brunfaut, T. (2014). A lifetime of language testing: An interview with J. Charles Alderson. *Language Assessment Quarterly*, 11(1), 103-119.



- Butler, Y. G., & Lee, J. (2010). The effects of self-assessment among young learners of English. *Language Testing*, 27(1), 5–31.
- Chapelle, C. A. (2006). Test review. *Language Testing*, 23(4), 544-550.
- Chen, J., Belkada, S., & Okamoto, T. (2004). How a Web-based course facilitates acquisition of English for academic purposes. *Language Learning & Technology*, 8(2), 33-49.
- Chen, Y. M. (2006). Peer and self-assessment for English oral performance: A study of reliability and learning benefits. *English Teaching and Learning*, 30(4), 1-22.
- Chen, Y. M. (2008). Learning to self-assess oral performance in English: A longitudinal case study. *Language Teaching Research*, 12(2), 235-262.
- Council of Europe. (2001). *The common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Cram, B. (1995). Self-assessment: From theory to practice. In: G. Brindley (Ed.), *Language assessment in action: Developing a workshop guide for teachers* (pp.271–305). NCELTR: Sydney, NSW.
- Dann, R. (2002). *Promoting assessment as learning: Improving the learning process*. New York: Routledge.
- Davies, A. (2003). Three heresies of language testing research. *Language Testing*, 20(4), 355-368.
- Deakin-Crick et al. (2005). Research evidence of the impact on students of self-and peer-assessment. EPPI-Centre: London.



- Escribano, I. D., & McMahon, J. P. (2010). Self-assessment based on language learning outcomes: A study with first year Engineering students. *Revista Alicantina de Estudios Ingleses*, 23, 133-148.
- Esfandiari, R., & Myford, C. M. (2013). Severity differences among self-assessors, peer-assessors, and teacher assessors rating EFL essays. *Assessing Writing*, 18, 111–131.
- Falchikov, N. (1986). Product comparisons and process benefits of collaborative peer group and self-assessments. *Assessment and Evaluation in Higher Education*, 11(2), 146–165.
- Haahr, J. H., & Hansen, M. E. (2006). *Adult skills assessment in Europe: Feasibility study*. Danish Technological Institute.
- Hamp-Lyons, L. (2007). Final report of the longitudinal study on the school-based assessment component of the 2007 HKCE English language examination. Report submitted to the Hong Kong Examinations and Assessment Authority, November.
- Hana Lim, H. (2007). A Study of self- and peer-assessment of learners' oral proficiency. *CamLing*, 169-176.
- Hubbard, P., & Levy, M. (Eds.). (2006). The scope of CALL education. In P. Hubbard & M. Levy (eds.), *Teacher education on CALL* (pp. 3-20). Amsterdam: John Benjamins Publishing Company.
- Kato, F. (2009). Student preferences: Goal-setting and self-assessment activities in a tertiary education environment. *Language Teaching Research*, 13(2), 177-199.
- Klimova, B. F., & Hubackova, S. (2013). Diagnosing students' language knowledge and skills. *Procedia - Social and Behavioral Sciences*, 82, 436-439.



- LeBlanc, R., & Painchaud, G. (1985). Self-assessment as a second language placement instrument. *TESOL Quarterly*, 19(4), 673-687.
- McDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: the effects of self-assessment training on performance in external examinations. *Assessment in Education*, 10(2), 209–220.
- Moritz, Ch. E.B. (1996). Student self-Assessment of language proficiency: Perceptions of self and others. A paper presented at AAAL Chicago Conference. Retrieved from Eric.
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2): 199–218.
- North, B. (2000). Defining a flexible common measurement scale: Descriptors for self and teacher assessment. In: G. Ekbatani & H. Pierson (Eds.), *Learner-directed assessment in ESL* (pp.13-48). Mahwah, NJ: Lawrence Erlbaum Associates.
- Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment and Evaluation in Higher Education*, 25(1), 23-38.
- Orsmond, P., Merry, S., & Reiling, K. (2002). The use of exemplars and formative feedback when using student derived marking criteria in peer and self-assessment. *Assessment and Evaluation in Higher Education*, 27(4), 309–323.
- Oscarson M (1989). Self-assessment of language proficiency: Rationale and applications. *Language Testing*, 6(1), 1–13.





- Oscarson, M. (1997). Self-assessment of foreign and second language proficiency. In *The encyclopedia of language and education, Vol. 7: Language testing and assessment* (pp. 175–187). Dordrecht: Kluwer Academic.
- Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychology, 36*(2), 89–101.
- Peden, B. F., & Carroll, D. W. (2008). Ways of writing: Linguistic analysis of self-assessment and traditional assignments. *Teaching of Psychology, 35*(4), 313–318.
- Pierce, B. N. et al. (1993). Self-assessment, French immersion, and locus of control. *Applied Linguistics, 14*(1), 25-42.
- Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experimental factors. *Language Testing, 15*(1), 1–19.
- Sadler, P., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment, 11*(1), 1-31.
- Saito, H., & Fujita, T. (2004). Characteristics and user acceptance of peer rating in EFL writing classrooms. *Language Teaching Research, 8*(1), 31–54.
- Sally, A. (2005). How effective is self-Assessment in writing? In P. Davidson, C. Coombe, & W. Jones (Eds.), *Assessment in the Arab world*. (pp. 307-321). United Arab Emirates: TESOL Arabia.
- Segers, M., & Dochy, F. (2001). New assessment forms in problem-based learning: The value added of the students' perspective. *Studies in Higher Education, 26*(3), 327–343.



Shohamy, E. (2001). Democratic assessment as an alternative. *Language Testing*, 18(4), 373–391.

Silye, M. F., & Wiwczarowski, T. B. (2002). A critical review of selected computer assisted language testing instruments. [online] University of Debrecen, Centre of Agricultural Sciences, Faculty of Agricultural Sciences, Centre of Technical Languages Instruction, Debrecen. Retrieved from <http://www.date.hu/acta-agraria/2002-01i/fekete1.pdf>

Son, J.-B. (2008). Using Web-based language learning activities. *International Journal of Pedagogies and Learning*, 4(4), 34-43.

Stanley, J. (1992). Coaching student writers to be effective peer evaluators. *Journal of Second Language Writing*, 1, 217–233.

Stefani, L. (1998). Assessment in partnership with learners. *Assessment and Evaluation in Higher Education*, 23(4), 339–350.

Stiggins, R. J. (2001). *Student-involved classroom assessment* (3rd ed.). Upper Saddle River, NJ: Merrill.

Suzuki, M. (2009). The compatibility of L2 learners' assessment of self- and peer revisions of writing with teachers' assessment. *TESOL Quarterly*, 43(1), 137-148.

Taras M (2001). The use of tutor feedback and student self-assessment in summative assessment tasks: Towards transparency for students and for tutors. *Assessment and Evaluation in Higher Education*, 26(6), 605–614.

Taras, M. (2003) To feedback or not to feedback in student self-assessment. *Assessment and Evaluation in Higher Education*, 28(5), 549 – 565.



- Todd, R.W. (2002). Using self-assessment for evaluation. *English Teaching Forum*, 40(1), 16–19.
- Topping, K. J. (2003). Self-and peer-assessment in school and university: Reliability, validity and utility. In: M. Segers & E. Cascallar (Eds.), *Optimizing new methods of assessment: In search of qualities and standards* (pp. 55–87). Dordrecht, Netherlands: Kluwer Academic Publishers.
- Towler, L., & Broadfoot, P. (1992). Self-assessment in the primary school. *Educational Review*, 44(2), 137-151.
- Wagner, L., & Lilly, D. H. (1999). Asking the experts: Engaging students in self-assessment and goal setting through the use of portfolios. *Assessment for Effective Intervention*, 25(1), 31-43.
- Warschauer, M., & Meskill, C. (2000). Technology and second language learning. In J. Rosenthal (Ed.), *Handbook of undergraduate second language education* (pp. 303-318). Mahwah, New Jersey: Lawrence Erlbaum.
- Wiggins, G. P. (1993). *Assessing student performance: Exploring the purpose and limits of testing*. San Francisco: Jossey Bass.



### **Authors**

**Mahboubeh Taghizadeh** holds a PhD in TEFL from University of Tehran, an MA in TEFL from Iran University of Science and Technology, and a BA in English Language Literature from Az-Zahra University. She has published in some national and international journals. Some of her current interests include CALL, language assessment, ESP, and teacher education. E-mail: mah\_taghizadeh@ut.ac.ir

**Sayed Mohammad Alavi**, an associate professor of Applied Linguistics in University of Tehran, is interested in doing research in language testing, English for specific purposes, CALL, and materials development. He teaches the courses related to his research interests to BA, MA, and PhD students. E-mail: smalavi@ut.ac.ir

**Abbas Ali Rezaee** obtained his PhD in Applied Linguistics from the University of Exeter in England. He is currently an Associate Professor in the Department of English Language and Literature at the Faculty of Foreign Languages and Literatures, University of Tehran. He has published extensively in national and international journals. He has taught various courses in language teaching and testing at BA, MA and PhD levels for more than 25 years. His main research interests are English Language Teaching, Language Testing, CALL, and Issues in ESP. E-mail: aarezaee@ut.ac.ir



This work is licensed under a [Creative Commons Attribution 3.0 License](https://creativecommons.org/licenses/by/3.0/).

