Comparing Male and Female Engineering Students' Reading Ability through Level-Specific Tasks and CEFR- and DIALANG-Based Self-Assessment

Mahboubeh Taghizadeh[1], Farshad Kazemzadeh[2]

ARTICLE INFO

ABSTRACT

The purposes of this study were to determine a reading ability level based on the CEFR for male and female engineering students and to compare their reading levels on the CEFR and DIALANG self-assessment grids with the level based on their performance on the reading tasks. The participants were 162 Iranian male and female undergraduate students at the Iran University of Science and Technology. DIALANG and CEFR reading self-assessment grids and four reading tasks at four levels (i.e., A2, B1, B2, and C1) were administered to the participants. Considering the students' performance on the reading tasks, the results revealed that female learners outperformed the male learners; however, a considerable number of learners (about 40%) in both groups were considered 'Below B1'. The results also showed that learners' self-assessment did not closely correspond with their performance on the reading tasks as both groups of learners rated their reading ability lower on the CEFR and DIALANG grids; nevertheless, the highest matches were related to the female learners' ratings on the DIALANG grid.

KEYWORDS

---

[1] Assistant Professor, Iran University of Science and Technology, mah_taghizadeh@iust.ac.ir
[2] MSc, Iran University of Science and Technology

Introduction

In the 1990s, some educators started questioning the validity of measuring language knowledge by traditional tests as they require test takers to attempt tests, which are based on artificial and contrived language content (Salmani Nodoushan, 2008) As the problems and shortcomings of standardized testing surfaced, the term 'alternative assessment' came into view in this period (Brown, Irving, Peterson, & Hirschfeld, 2009). Brown and Abeywickrama (2004) assert that alternative assessments use authentic tasks, which indicate what learners can do with language in real life and educational contexts. Brown and Abeywickrama (2004) also note that these tests are different in design and structure from traditional tests and are graded differently as well. Alternative assessment, as Huerta-Macias (2002) asserts, consists of portfolio, journal, observation, self-assessment, peer-assessment, interview, and conference, which can give a holistic view of learners' abilities, shed light on the parts in which learners need further improvement, and give the students the chance to demonstrate the scope of learning (El-Henawy, 2017).

It is argued that students can learn better when they are encouraged to think about what they are learning and how they learn it (McDonald & Boud, 2003). Self-Assessment (SA) is one of the most important types of formative assessment, which is a procedure in which learners evaluate their language knowledge and skills and can enhance learners' motivation and encourage them to seek their weaknesses when participating in a self-assessment (Brown & Harris, 2013). Self-assessment is defined as a process in which the learner assesses the actions he or she engages in while learning a particular subject or skill and is usually conducted with the intention of identifying one's strengths and weaknesses and improving the learning outcomes (Krawiec, 2014). Self-assessment, according to Pierce and Durán (2014), can make the hidden processes more concrete and visible, allowing learners to develop their inner measurement of progress.

The reading self-assessment statements of DIALANG and CEFR projects are used in this study. CEFR is designed to prepare a coherent and comprehensive basis for the language curriculum guideline, language syllabuses, and assessment of language proficiency (Council of Europe, 2001) CEFR represents six levels of language proficiency (i.e., A1, A2, B1, B2, C1, and C2), which enable us to compare tests across languages and national boundaries (Morrow,

2004) and are validated in quantitative and qualitative studies (Alderson, 2002; Hasselgreen, 2003). The statements of the levels are skill-based and are in the form of *can do* statements. These statements focus on the communicative aspects, which maintain what people can do instead of knowledge of separate linguistic items. It is also useful for teachers in understanding the proficiency level of each learner, identifying more precisely the areas in which learners need further work, using the CEFR scales in creating their own grids, and making the curriculum plan (Council of Europe, 2001).

The reason for worldwide attention to the CEFR is the fact that it offers a more comprehensive and detailed system of level descriptions than the other systems and that it was developed on the basis of research in second language acquisition, foreign language education, and test research (Bärenfänger & Tschirner, 2008). Furthermore, the framework is being used as a set of guidelines in order to describe the learners' achievements of foreign languages and as a reference document to promote understanding and management about the various levels of the educational system by providing a reference for language skills (Gouveia, 2006). Doing an experimental validation through the L-scale algorithm, Lindhout, Teunissen, and Lindhout (2012) calculated an accurate effective readability level for CEFR.

DIALANG's assessment framework and the descriptive scales used for reporting the results to users are directly based on the CEFR (Alderson, 2005). It is an assessment tool prepared for those who have learnt a language and want to know their proficiency level. The self-assessment statements used in the DIALANG are also mostly based on the CEFR and adapted whenever essential to fit the specific needs of the system (Council of Europe, 2001). Two purposes of self-assessment in DIALANG are to encourage autonomy in learning and to use self-assessment results to pre-estimate learners' ability (Morrow, 2004).

As Weir (2005) states, the initial development of the self-assessment section of DIALANG was conducted by a self-assessment working group, who started the development from the scale descriptors of CEFR. The DIALANG self-assessment working group simplified the language in the statements a little and changed the formulations from 'can do' to 'I can'. In addition, a small number of new statements had to be created because the statements in the CEFR did not fully cover all aspects and levels of proficiency (Milanovic & Weir, 2004). Weir

also asserts that the criteria for selecting the statements should be concrete and practical enough for learners to understand, fit the general purpose orientation of the DIALANG test sections, and be highly advanced. There are self-assessment statements for reading, writing, and listening, but not for vocabulary and grammar as the CEFR does not include any language-specific self-assessment statements for these two components (Goodier, 2014). Alderson (2002) states that the statements are assembled in the order of difficulty in which learners have to read each statement in turn and simply respond yes or no. Alderson also argues that once users have taken the language test properly, the result of the self-assessment is reported and compared with the user's test result.

Researchers have shown interest in investigating reading as one of the major problems in reading assessment is the process of reading, which is abstract, internal, and hidden (Duncan, 2012). Reading is among the skills, which have indicated the most challenges for language learners (Grabe, 2009) and thus has been investigated from various aspects; however, to the researchers' best knowledge, no study has been conducted estimating Engineering students' reading ability through level-specific CEFR-based tasks and self-assessment grids. Accordingly, the purposes of this study were to determine a reading level based on CEFR for male and female undergraduate students of engineering and to compare their levels in terms of the CEFR and DIALANG reading self-assessment grids with the level specified based on their performance on some reading tasks. The research questions formulated for this study were as follows:

1. Is there any difference between male and female Engineering students in their reading ability in terms of the level-specific tasks based on CEFR?
2. Is there any difference between learners' self-assessed level of reading ability in terms of CEFR statements and the level specified based on their performance on the reading tasks?
3. Is there any difference between learners' self-assessed level of reading ability in terms of DIALANG statements and the level specified based on their responses to the reading tasks?

Method

Participants

This research was conducted with 162 Iranian male and female undergraduate students (male=78, female=84) at the Iran University of Science and Technology (IUST). The participants' majors were mechanical engineering (*n*=19), electrical engineering (*n*=20), railway engineering (*n*=21), material engineering (*n*=16), industrial engineering (*n*=19), chemical engineering (*n*=21), industrial design (*n*=9), computer engineering (*n*=15), architecture (*n*=11), and civil engineering (*n*=11). They were all adult learners ranging in age from 18 to 22.

Instruments and Materials

In this research, to facilitate learners' better understanding, the translated versions of the DIALANG and CEFR reading self-assessment grids (see Appendices A & B) were administered to the participants. The DIALANG reading self-assessment grid consisted of 31 statements, while the CEFR scale was comprised of 11 self-assessment statements for the reading skill. All the statements used in the questionnaires were binary with yes or no responses. The detailed information about the different levels of each scale is presented in Table 1.

Table 1: *The Number of Self-Assessment Statements for the Six Levels of the DIALANG and CEFR Scales*

Scale	Levels						
	A1	A2	B1	B2	C1	C2	Total
CEFR	2	2	2	2	2	1	11
DIALANG	6	8	8	6	2	1	31

In order to collect the needed data, four reading comprehension tasks at four levels (i.e., A2, B1, B2, and C1) were also administered. The first task, 'Learn English in Christchurch' was chosen from *Real Reading 1* (Driscoll, 2008) in which the learners were asked to read two sections from a school webpage and then fill in the blanks with the appropriate words. The second task titled 'Join a Library' was chosen from *Real Reading 2* (Driscoll, 2008) in which the learners were requested to read a text from a website and decide whether some statements were true or false. The third task titled 'Kinds of Reading' was chosen from *Real Reading 3* (Driscoll, 2008) in which the learners were asked to circle the correct phrases. The

final task, 'Obstacles to Faster Effective Reading', was chosen from *Real Reading 4* (Driscoll, 2008) in which the learners were requested to decide whether some statements were true or false according to the information given in the passage.

Procedure

The following procedure was used in conducting this research. Before administering the tasks to the participants of the study, they were informed of the purposes of the research. DIALANG and CEFR self-assessment grids were translated to Persian and the translated versions were then revised by four experts who had an MA or a PhD in TEFL. This was done not only to have explicit and unambiguous grids, but also to assess their quality before they were used with the actual participants. Subsequently, the grids were piloted with 13 students in order to estimate the probable, administrative problems and also to calculate the approximate working time.

The reading tasks were then chosen from the *Real Reading* series, which were designed and published based on the A2, B1, B2 and C1 levels of the CEFR. Tasks were also checked by four PhD holders in TEFL who had specific expertise in language testing. Tasks were piloted with 13 learners so as to eliminate ambiguities, to check on the clarity and comprehensibility of the questions and the rubrics, to have the first impression of the difficulty level of the tasks and items, to estimate the time load involved, and to determine the possible problems in the administration of the tasks. The validation process for the tasks used in this research was a qualitative validation performed by the authors of the *Real Reading* textbooks in which they asked a panel of recognized experts to review their developed reading tasks.

In this study, the researchers omitted A1 and C2 from the six levels of the CEFR as it was assumed that Iranian undergraduate students already reached the A1 level as they passed many reading-based language courses before entering university. The reason for the removal of the C2 level was related to the fact that no task was found to test this level as mostly only natives could reach this level. The validation procedures for the tasks used in this study were also based on what the authors of the *Real Reading* textbooks reported for their developed tasks. It is worth noting that the researchers chose the tasks which closely

matched the sample and excluded the tasks affected by cultural biases or the ones which might have been recognized as offensive by Iranian students. Next, the participants were asked to answer 31 self-assessment statements of DIALANG and 11 self-assessment statements of CEFR with yes/no responses. Learners were also asked to perform four tasks in 45 minutes (i.e., 5 minutes for the A2 task, 10 minutes for the B1, and 30 minutes for the B2 and C1 tasks).

The detailed analyses of the learners' performance on the reading tasks formed the basis for a final overall reading score. Additionally, the researchers collected all the self-assessment scores for both DIALANG and CEFR grids in order to determine one CEFR self-assessed level per person. In other words, each self-assessment statement was scored with 1, and it was assumed that, for instance, a C1 learner was expected to 'solve' all A2, B1, and B2 statements plus 75% of the C1 statements.

To answer the research questions addressed in this study, the following statistical analyses were performed. Descriptive statistics were conducted to determine each learner's reading level based on their performance on the four adjacent reading tasks. In addition, descriptive statistics were performed for the self-assessment statements of the CEFR and DIALANG scales. Finally, descriptive statistics for the three categories of underrate, overrate, and match for the self-assessment grids were also calculated.

Result

Self-Assessment Statements of the DIALANG Scale

In this section, the percentage for each reading self-assessment statement of the DIALANG grid for both female and male learners is presented.

In order to determine which items received more positive replies, the percentage of female students' responses to each item of the DIALANG scale was calculated. The results showed that, item 4, *'I can recognize familiar names, words and very simple phrases on simple notices in the most common everyday situations'* and item 5, *'I can understand short, simple messages, e.g. on postcards'*, both gained the highest percentage (97.6%), while item 31, *'I can understand and interpret practically all forms of written language including abstract, structurally complex,*

or highly colloquial literary and non-literary writings', received the lowest percentage (9.5%). The frequency and percentage of male students' responses concerning each item of the DIALANG scale were also calculated. The results indicated that item 3, *'I can follow short, simple written instructions, especially if they contain pictures'*, received the highest percentage (97.4%), while item 31, '*I can understand and interpret practically all forms of written language including abstract, structurally complex, or highly colloquial literary and non-literary writings'*, received the lowest percentage (10.3%).

Self-Assessment Statements of the CEFR Grid

In this section, the frequency and percentage for each reading self-assessment statement of the CEFR grid for both female and male students are presented.

**Table 2: *Frequency & Percentage of Female Students' Replies to* **CEFR Self-Assessment Grid (N =86)

Levels	Items	f	%	Rank
A1	#1	81	96.4	2
A2	#2	82	97.6	1
A2	#3	62	73.8	6
A2	#4	73	86.9	3
B1	#5	71	84.5	4
B1	#6	63	75.0	5
B2	#7	24	28.6	7
B2	#8	18	21.4	8
B2	#9	7	8.3	10
C1	#10	8	9.5	9
C1	#11	5	6.0	11

Considering the reading descriptors of the CEFR grid for female students, as shown in Table 2, item 2, '*I can read very short, simple texts'*, obtained the highest percentage (97.6%), whereas item 11, '*I can read with ease virtually all forms of the written language, including*

abstract, structurally or linguistically complex texts such as manuals, specialized articles and literary works' received the lowest percentage (6.0%).

Table 3: *Frequency & Percentage of Male Students' Replies to **CEFR Self-Assessment Grid (N =78)***

Levels	Items	F	%	Rank
A1	#1	77	98.7	1
A2	#2	77	98.7	1
A2	#3	67	85.9	3
A2	#4	63	80.8	4
B1	#5	68	87.2	2
B1	#6	53	67.9	5
B2	#7	35	44.9	6
B2	#8	23	29.5	7
B2	#9	7	9.0	9
C1	#10	8	10.3	8
C1	#11	8	10.3	8

With regard to male students' performance on the CEFR grid, as indicated in Table 3**,** item 1, '*I can understand familiar names, words and very simple sentences, for example on notices and posters or in catalogs'* and item 2, '*I can read very short, simple texts'* received the highest percentage (98.7%)**,** whereas item 9, *'I can understand long and complex factual and literary texts, appreciating distinctions of style*' received the lowest percentage **(9%).**

Table 4: *Students' Reading Levels in Terms of their Performance, DIALANG and CEFR Grids in %*
Students' Performance on the Reading Tasks, CEFR and the DIALANG Grids

In this section, the results of the percentage and chi-square analysis for both female and male students in terms of the six levels of CEFR are presented.

As shown in Table 4, female students performed better than male students on the A1 and C1 tasks. About overall, 33.3% of female students were placed at C1, while the majority of

male students (30.8%) were placed at B2. Table 4 also shows that the majority of female students were at the C1 and A1 levels, whereas male students were at the B2 and A1 levels. B2 was the level at which female learners showed the weakest performance (9.5%), while B1 and C1 were the levels at which male learners showed the weakest performance (15.4%). Table 4 also indicates that approximately a considerable number of participants (about 40%) in both groups were considered 'Below B1'.

As indicated in Table 4, 40.5% of female students rated themselves at B1, whereas just only 8.3% assessed their reading ability at B2 and C2 on the DIALANG grid. On the other hand, with regard to the CEFR grid they assessed themselves mostly at B1 (48%), while C2 was their least reported level. It is important to note that no female student rated herself at C1 on both the CEFR and DIALANG grids. The most frequent self-assessed level for male students was B1 on the DIALANG, while the least frequent level was C1. About one-third of them (37.2%) assessed themselves at B1, whereas only 9% at C2 on the CEFR grid. Participants' performance on the reading tasks, DIALANG, and the CEFR grids could be hierarchically ranked as: Female: (*rated performance*: C1, A1, A2, B1, B2; *DIALANG*: B1, A1, A2, B2 , C2, C1; *CEFR*: B1, A2, B2, A1, C2, C1); Male: (*rated performance*: B2, A1, A2, B1, C1; *DIALANG*: B1, A1, A2 , B2, C2, C1; *CEFR*: B1, A2, B2, A1, C2, C1). In general, participants assessed themselves mostly at B1 on both DIALANG and CEFR grids; however, their performance on reading tasks was significantly different in which females were placed at C1, while male students were placed at B2.

Underrate, Overrate, and Match Categories on the DIALANG and CEFR Grids

Table 5: *Percentage of Underrate, Overrate, and Match Categories on DIALANG and CEFR*

Gender	SA Grids	Underrate	Overrate	Match
Female	DIALANG	46.4	25.0	28.6
	CEFR	42.9	29.8	27.4
Male	DIALANG	51.3	23.1	25.6
	CEFR	48.7	32.1	19.2

Grids for Female and Male Language Learners

As indicated in Table 5, both male and female respondents rated their reading ability lower on both CEFR and DIALANG grids. The highest matches (28.6%) were related to the female students' ratings on the DIALANG grid, while the highest mismatch (51.3%) was the underestimation of the male students on the DIALANG grid. Table 5 also shows that both female and male students considerably underrated both on the DIALANG and CEFR grids. About half of the female students tended to underestimate on the DIALANG, while 25% tended to overrate and about one-third matched. The majority (42.9%) of the female students underrated on the CEFR, while approximately one-third overrated and matched on the CEFR. Table 5 also indicates that male students showed the highest underrated mismatch on the DIALANG and CEFR grids. However, male students received the lower percentage for the match category than the female students did.

Discussion

The performance of both male and female students on the reading tasks was not satisfactory. The major goal of English language courses offered in Iran is to improve students' reading ability in order to prepare them for the scientific texts in universities; however, English classes in schools and even in universities mostly turn to translation classes (Chalak, 2015) where students are asked to memorize the meanings of words and translate the sentences into Persian.

In Iran, teaching English starts in the 6th grade, when learners have lost the best years for learning a foreign language. The delay in foreign language learning in the public education seems to be one of the important language learning problems leading to poor ability to

comprehend and communicate the language. In addition, despite the presentation of grammatical and lexical points to students, the results are not favorable, which might be due to the fact that learners only try to memorize the new concepts and meanings, and as a result after a few days they forget what they have learned.

As another reason for Iranian students' poor reading comprehension, unattractive textbooks can be stated (Ahymadpoor, 2004) The books taught in our educational system are not up-to-date and include colorless pictures and pages, and the texts not in line with the real life tasks. Moreover, the lack of knowledge about the strategies for learning new words is another major factor for learners' poor comprehension (Ahmadi & Mahmoodi, 2012). In addition, heterogeneity of students' level of language proficiency (Sadeghi & Richards, 2016) has a direct impact on the learning and comprehending English reading texts. It is also believed that the lack of standards for the reading proficiency and the lack of predetermined, concrete reading outcomes, the traditional teacher-centered teaching methods and time constraints can also account for the problems in the reading ability of the learners. Additionally, according to (Nezakatgoo & Behzadpoor, 2017) and the great number of students in the classes in Iran might be another reason as all students cannot participate in the class activities and cannot effectively express their ideas about their learning in the class.

It was found that most students of this research did not have clear and accurate perceptions of their reading ability as their self-assessments did not correspond closely with their actual performance. This might be due to the fact that in the educational system of Iran learners are not often asked to assess themselves or their reading abilities, and their assessment is mostly summative in which the results of their performance are reported as a single score.

The results also showed that the number of matches in both male and female performance on the DIALANG was more than that on the CEFR grid. This may in part be due to the fact that the number of reading statements for each level on the DIALANG grid is more than that on the CEFR. In other words, it can be claimed that DIALANG could effectively and clearly illustrate the students' ability.

The findings of this study are in line with those of Ashton (2014) in that students achieved lower scores on the reading tasks (i.e., report reading) and also learners

underestimated themselves on the self-assessment grids. In another study, Ünaldı (2016) used self and teacher assessments in order to predict the proficiency level of the first year students of university in which the results demonstrated that participants underestimated themselves on the CEFR and DIALANG grids which are in line with the results of the present study.

Conclusions

This study aimed to determine a reading level based on CEFR for the undergraduate students of engineering at the IUST and to compare their reading self-assessment on the CEFR and DIALANG grids with their actual reading level. Four reading tasks and the DIALANG and CEFR reading self-assessment grids were administered to the participants. The results showed that (a) the majority of both male and female participants rated their reading ability lower both on the CEFR and DIALANG grids; (b) the learners' self-assessment did not correspond closely with their performance on the reading tasks, and approximately only one-third of them in both male and female groups were accurate in assessing their reading, while about one-third of them tended to overrate their reading performance; (c) the highest matches for the male students were related to their ratings on the DIALANG grid, while the highest mismatch was the underestimation on the DIALANG grid; (d) the highest matches for female group were achieved on the DIALANG grid, while the highest mismatch was their underestimation on the DIALANG grid; and (e) both female and male students rated themselves at B1 level on the CEFR and DIALANG grids. Regarding their performance on the reading tasks, most female students were placed at the C1 level, while most male students were at the B2 level.

Incorporation of self-assessment in reading instruction can promote learners' sense of independence in language learning and help them overcome their fears in reading. Students can also use *can do* statements to evaluate their progress in reading and to formulate certain goals for their future progress. It is believed that there is the lack of standard teaching and assessment procedures for reading skill at the secondary and tertiary levels; thus, language teachers are suggested to use CEFR descriptors for teaching and assessing reading skill.

Among different types of alternative assessment, self-assessment was utilized in this study in order to estimate the reading performance of undergraduate students of

Engineering. Interested researchers can investigate the effect of other types of alternative assessment such as peer assessment, portfolio, and conference on EFL learners' reading skill. Researchers in future can also use in-depth interview with learners and instructors about learners' reading performance with respect to *can do* statements. The relationship between self-assessment in terms of CEFR and DIALANG statements and psychological factors such as personality traits, learning anxiety, and cognitive styles advantages further inquiry.

REFERENCES

Ahmadi, A., & Mahmoodi, S. (2012). Language learning strategy use and instruction for the Iranian Junior high school EFL learners: A mixed methods approach. *Research in Applied Linguistics, 3*(2), 107-134.

Ahymadpoor, Z. (2004). Studying the Problems of EFL Teaching in High Schools. *The Roshd ELT Journal, 18*(71), 15-21.

Alderson. (2002). *Case studies in the use of the Common European Framework*.

Alderson. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*: A&C Black.

Ashton, K. (2014). Using self-assessment to compare learners' reading proficiency in a multilingual assessment framework. *System, 42*, 105-119.

Bärenfänger, O., & Tschirner, E. (2008). Language educational policy and language learning quality management: The Common European Framework of Reference. *Foreign Language Annals, 41*(1), 81-101.

Brown, H. D., & Abeywickrama, P. (2004). Language assessment. *Principles and Classroom Practices. White Plains, NY: Pearson Education*.

Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. H. Mcmillan (Ed.), *The SAGE handbook of research on classroom assessment* (pp. 367–393). Los Angeles, CA: SAGE.

Brown, G. T., Irving, S. E., Peterson, E. R., & Hirschfeld, G. H. (2009). Use of interactive–informal assessment practices: New Zealand secondary students' conceptions of assessment. *Learning and Instruction, 19*(2), 97-111.

Chalak, A. (2015). The effect of task-based instruction on reading comprehension of Iranian EFL learners. *Applied Research on English Language, 4*(1), 19-30.

Council of Europe, (2001). *Common European Framework of Reference for Languages: learning, teaching, assessment*: Cambridge University Press.

Driscoll, L. (2008). *Real reading 1*. Cambridge: Cambridge University Press.

Driscoll, L. (2008). *Real reading 2*. Cambridge: Cambridge University Press.

Driscoll, L. (2008). *Real reading 3*. Cambridge: Cambridge University Press.

Driscoll, L. (2008). *Real reading 4*. Cambridge: Cambridge University Press.

Duncan, S. (2012). *Reading circles, novels and adult reading development*: Bloomsbury Publishing.

El-Henawy, W. M. (2017). Assessment Techniques in EFL Brain-Compatible Classroom *Innovative Practices for Higher Education Assessment and Measurement* (pp. 79-100): IGI Global.

Goodier, T. (2014). *Working with CEFR can-do statements: An investigation of UK English language teacher beliefs and published materials.* London.

Gouveia, C. A. (2006). The role of a common European framework in the elaboration of national language curricula and syllabuses. *Papers on Language and Society, 8*, 8-25.

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*: Ernst Klett Sprachen.

Hasselgreen, A. (2003). *Bergen "can do" project*. Strasbourg, France: Council of Europe.

Huerta-Macias, A. (2002). Alternative assessment: Responses to commonly asked questions *Methodology in language teaching* (pp. 338-343). Cambridge Cambridge University Press.

Krawiec, M. (2014). *Cross-Curricular Dimensions of Language Learning and Teaching*. Cambridge: Cambridge Scholars Publishing.

Lindhout, P., Teunissen, G., & Lindhout, M. (2012). Quick effective CEFR level evaluation of complete documents: Integrated readability appraisal of text and graphics with the L-scale algorithm. *International Journal of Language Studies, 6*(3), 37-56.

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: Principles, Policy & Practice, 10*(2), 209-220.

Milanovic, M., & Weir, C. (2004). *European language testing in a global context*. Cambridge: Cambridge University Press.

Morrow, K. (2004). *Insights from the common European framework*: Oxford University Press.

Nezakatgoo, B., & Behzadpoor, F. (2017). Challenges in Teaching ESP at Medical Universities of Iran from ESP Stakeholders' Perspectives. *Iranian Journal of Applied Language Studies, 9*(2).

Pierce, J., & Durán, P. (2014). Developing and Evaluating Self Assessment Reading Competence Descriptors for the European Language Portfolio. *Porta Linguarum, 22*, 127-143.

Sadeghi, K., & Richards, J. C. (2016). The idea of English in Iran: An example from Urmia. *Journal of Multilingual and Multicultural Development, 37*(4), 419-434.

Salmani Nodoushan, M. A. (2008). Performance Assessment in Language Testing. *i-manager's Journal on School Educational Technology, 3*(4), 1-7.

Ünaldı, İ. (2016). Self and teacher assessment as predictors of proficiency levels of Turkish EFL learners. *Assessment & Evaluation in Higher Education*, 41(1), 67-80.

Weir, C. J. (2005). *Language testing and validation*. Houndgrave, Hampshire, UK: Palgrave McMillan.