International Journal of Research in English Education

Online ISSN: 2538-3027

Homepage: https://ijreeonline.com

EFL Students' Rating Accuracy in Assessing Reading Comprehension Subskills across Genres: A Diagnostic Perspective

Masoome Azmoode[1*]

1. Department of foreign Languages, Science and Research University, Tehran, Iran, Email: azmoode2014@gmail.com

Article Info

Article type:

Research Article

Article history:
Received: 25 May 2025
Revision: 17 September 2025
Accepted: 21 September 2025
Published online: 30 September 2025

***Keywords*:**
reading comprehension,
reading subskills,
reading genres,
rating accuracy,
diagnostic assessment

ABSTRACT

This study aimed to examine EFL learners' rating accuracy in assessing reading comprehension subskills in various genres. To this end, 60 English translation students participated in this study. The instructional treatment was based on the learners' challenging reading subskills in four genres. Taking the instructor's ratings as the yardstick, during a 12- week course, the accuracy of the learners' self- and peer- assessments was investigated. Data analysis, using MANOVA, confirmed that there was a statistically significant difference between the accuracy of self-, peer- and instructor-ratings. More specifically, the two groups were inaccurate in assessing the main idea/ supporting details and cause/effect subskills in all genres. However, for assessing fact /opinion subskills only the self-assessment group was inaccurate in descriptive genre. Obtaining such detailed diagnostic information about learners' performance can help instructors in elevating weaknesses in language skills.

Keywords: reading comprehension; reading subskills; reading genres; rating accuracy; diagnostic assessment

Cite this article: Azmoode M. (2025). EFL Students' Rating Accuracy in Assessing Reading Comprehension Subskills across Genres: A Diagnostic Perspective. *International Journal of Research in English Education*, *10*(3), 1-15.

CC BY NC

Publisher: Science Academy Publications.

[Downloaded from mail.ijreeonline.com on 2026-11-18]

1. Introduction

Using a foreign language, as a cognitive ability, can be viewed and assessed from two diverse perspectives: the "internal perspective" that entails the individual's own assessment, and the "external perspective" that includes an assessment made by someone else (Oscarson, 2013). Originating from such views, two modes of assessments namely, self-, and peer- assessment have become the focus of attention in instructional contexts. In fact, the profound changes in evaluation procedures and reorientation in the role of students in learning environments have put such nontraditional assessments or alternative assessments in sharp focus. Research on self- and peer-assessment stems from a broader field of inquiry referred to as "assessment for learning", it is mainly built upon a relation between teaching and learning with assessment (Pang, 2020). In the same vein, the theory of diagnostic assessment substantially emphasizes the role of learner involvement in diagnosis (Harding et al., 2015). Based on this view, various stakeholders' opinions need to be integrated into diagnostic decisions for obtaining a richer insight into particular learning difficulties (Alderson et al., 2015).

To what extent can students evaluate their own performance and that of their peers accurately? For finding an answer to this fundamental question, studies on whether the learners' rating matches with the teachers' rating (or any other reliable rating) became an interesting line of inquiry. Admittedly, the validity of self-, and peer assessment has always been an important concern in this field (Lee & Chang, 2005). Exploring the literature shows extensive body of research on reliability and validity of learners' assessment in general (Han, 2018; Ma & Winke, 2019), and on the accuracy of their assessment in particular (Birjandi & Siyyari, 2010; Han & Riazi, 2017; Lu, 2018). However, less research has been conducted on learner-assessment accuracy in receptive skills, like reading comprehension (Ashton, 2014; Paleczek et al., 2015).

For assessing reading, it is of paramount importance to see it as a multi-divisible skill or as a single global construct. Despite controversies, test developers often define reading in terms of several subskills and include particular test items to measure those subskills (Song, 2008). The literature predominantly suggests the divisible view of reading; however, still there isn't any consensus among the advocates on the number, type and scope of such skills (Karakoc, 2019). Despite a growing research interest in multi divisible view of reading, not only there isn't any agreement among researchers on the kind of subskills tested by different items, but also there isn't a specific hierarchy of difficulty among different reading subskills. Reaching to such understanding is mainly dependent on the text genres, text topic, readers' knowledge and purposes of reading (Harding et al., 2015). Among the related issues, the role of 'reading genres' has been largely overlooked in research on reading comprehension subskills in EFL contexts.

Therefore, this study chose to focus on EFL learners' rating accuracy in assessing reading comprehension subskills in texts from different genres. Research with such orientation can help teachers see the importance of raising students' awareness regarding various reading components and types in instruction.

2. Literature Review

2.1 Self-and Peer-assessment

Self- and peer-assessment have been referred as involving students in assessing their own performance and that of others (Black & Wiliam, 2009). The students' involvement in assessment can assist them to find a clear picture about success criteria and manage their own development (Green, 2018). Self- and peer-assessment which stems from assessment for learning, is related to some other fields including self-regulation (Bailey & Heritage, 2018), learner autonomy (Benson & Voller, 2014; Singh Negi & Laudari, 2022), self-efficacy (Schunk, 2004), metacognition (Oxford, 2016), and motivation (Dornyei & Ushioda, 2013). More specifically, learner-assessment is central in the student-centered approach. In this approach the assessment burden is on the learners' shoulders in order to facilitate

[Downloaded from mail.ijreeonline.com on 2026-11-18]

learning process (Santos & Semana, 2015). Through self- and peer-assessment, learners are actively engaged in setting learning goals, monitoring their improvement, and decision making in learning process (Esfahani et al., 2022). This is in line with Vygotsky's (1978) Sociocultural Theory (SCT), the application of which in instructional settings can assist learners become interactive, autonomous, and responsible. Within the sociocultural paradigm, learning is perceived as the process of involving in a community of practice with more expert members, where mastery is developed in both social and cognitive ways through making use of cultural tools. In SCT, assessment activities are viewed as cultural tools to define learning possibilities (Lave, 1993).

Reviewing the related literature shows that in the design of many studies which investigated the learners' performance through self- and peer- assessment (Chen, 2008; Han & Riazi, 2017; Lu, 2018; Ma & Winke, 2019) some diagnostic purposes are implied. Since, employing self-and peer- assessment techniques increases learners' awareness regarding their own difficulties in language skills. However, identifying learners' strengths and weaknesses through informing instructional strategies and personalized learning experiences is the most important mission of diagnostic assessment (Alderson et al., 2015). In fact, diagnostic assessment is linked with tailored assessments by focusing on students' problems in language learning (or use) along with the source of the problems to help teachers and learners deal with the root causes of the weaknesses effectively (Lee, 2015). One of the pivotal principles of diagnostic assessment theory is stake-holder involvement (including learners themselves). Looking at diagnostic studies which favored learner involvement in their design (Azmoode et al., 2024 a; Azmoode et al., 2024 b; Markey, 2020; Mazloomi & Khabiri, 2016; Ng, 2018) shows that engaging students in assessment practices is beneficial in learning.

Many researchers have investigated the accuracy and reliability of learners-assessment by considering teachers' ratings, objective tests, or final grades as the reference point (Suzuki, 2015). The results of these studies (Alibakhsh, 2013; Aminu et al., 2021; Nalbantoğlu, 2017; Ross, 2006; Xiao & Lucking, 2008) mainly advocate the validity of learners' assessments. Using criterion measures to decide on the accuracy of the learners' assessments sometimes leads to a considerable measurement error variation, however. Two main sources of inconsistency in learners' assessments have been reported as "students' individual characteristics" and "skill types or skill domain" (Suzuki, 2015). Regarding the effect of individual characteristics on self-, and peer- assessment, researchers worked on some issues such as, age-related differences (Butler, 2018), gender (Van Kraayenoord & Paris, 1997) , self-esteem (AlFallay, 2004) individuals' learning progressions (Goral & Bailey, 2019), experiences (Butler & Lee, 2006; Suzuki, 2015),and self-regulation (Bailey & Heritage, 2018).

With respect to the role of skill types in students' assessment , many scholars have examined the accuracy of learner- assessment in different language skills (Birjandi & Siyyari, 2010; Han & Riazi, 2017; Lu, 2018); nevertheless, one can find fewer studies on self- and peer- assessment accuracy in receptive skills, like reading (Ashton, 2014; Paleczek, Seifert, Schwab, & Gasteiger-Klicpera, 2015; Ross 1998). Ross (1998), for instance, proved that learners were more accurate at self-assessing receptive skills. The current research also contributes to the field by examining the accuracy of learners' assessments in assessing a number of reading comprehension subskills across various genres.

There has been a substantial research growth in the domain of comprehension subskills in recent years. However, the literature shows a contradictory position regarding the divisibility of reading construct since a long time: while some researchers view reading as an integrated skill, in favor of general-factor theories (Goodman, 1967; Vacca, 1980); others believe in multiple-factor theories and consider reading as a divisible skill (Davis, 1944; Spearritt, 1972). Although the literature mostly advocates the multi divisible view of reading (Elahi, 2016; Farhady & Daftarifard, 2006; Karakoc, 2019; Tengberg, 2018), there is not enough evidence for its psychometric divisibility (Tengberg, 2018). While a lot of researchers have classified reading into various dimensions, there is currently little agreement on how the processes of reading comprehension can be categorized validly; thus, one can find no agreed- upon taxonomy for classifying comprehension subskills in research (Aryadoust, 2020). Though, it has been suggested that for addressing

[Downloaded from mail.ijreeonline.com on 2026-11-18]

the divisibility of comprehension subskills, some factors such as L2 proficiency of test takers a long with tests characteristics in certain testing contexts should be accounted (Song, 2008).

To advocate the multidivisible view of language skills, different comprehension subskills have been investigated so far and various taxonomies have been suggested accordingly. Through a document co-citation analysis Elahi (2016) focused on the comprehension subskills which were studied empirically since 1945 (in both L1 and L2). The findings showed that in L1 literature comprehension subskills (mainly the lower- order ones) are considered as dynamic and process-oriented whereas in L2 studies, subskills are seen as static and product-oriented. In the domain of cognitive diagnostic assessment also, researchers found multiple comprehension subskills by applying diagnostic models (Aryadoust, 2019; Javidanmehr & Anani Sarab, 2019; Lee & Sawaki 2009; Ranjbaran & Alavi, 2017; Ravand, 2015).

Previous findings suggest that for gaining a deeper recognition about the nature of reading subskills the characteristics of "text genre" need to be considered (Harding et al., 2015; Jang, 2009). While previous studies (Azmoode et al 2024 a; Javidanmehr & Anani Sarab, 2019; Ravand, 2015; Rouhi et al., 2015) proved that some reading genres or subskills are difficult to learn and comprehend for learners, no agreed- upon hierarchy of difficulty has been reported yet. The difficulty that students experience in learning specific concepts might be related to their cognitive overload. In fact, cognitive load refers to the mental effort required to process information during learning process (Surbakti et al., 2024). According to Cognitive Load Theory, the capacity of working memory is limited in learners thus imposing too much cognitive load diminishes learning efficiency (Ou, 2022).

In line with this view, Yoshida (2012) asserts that different genres invoke different cognitive processes, the expository texts, for instance, apply individual item processing whereas narrative texts apply relational processing; thus, texts in the narrative genre are thought to be easier to comprehend. In narrative texts there are a lot of related concepts that cause key propositions to be rehearsed frequently for better comprehension. Despite significant research attention to the assessment of reading comprehension (Azmoode et al 2024b; Javidanmehr & Anani Sarab, 2019; Tengberg, 2018), insufficient genre and subskills differentiation in studies on rating accuracy is ostensible. Therefore, this paper aims to gain insights on EFL learners' rating accuracy in assessing reading comprehension subskills in texts from different genres. More specifically, it addresses the following question:

RQ**1:** Is there any statistically significant difference between the accuracy of self-, peer-, and instructor-ratings in assessing learners' reading comprehension in four different reading genres?

RQ 1.1: Is there any statistically significant difference between the accuracy of self-, and instructor-ratings in assessing learners' reading comprehension in four different reading genres?

RQ1.2**:** Is there any statistically significant difference between the accuracy of peer-, and instructor-ratings in assessing learners' reading comprehension in four different reading genres?

RQ1.3: Is there any statistically significant difference between the accuracy of self-, and peer-ratings in assessing learners' reading comprehension in four different reading genres?

3. Methodology

3.1 Participants

Sixty Iranian EFL learners (19 to 23 years old) who majored in English translation at Islamic Azad University took part in this study. The participants were taking a Reading Comprehension course in two intact classes. The selection of the participants was based on their performance on Oxford Placement Test (OPT) (M = ۳۳/۹۵, SD =۵/۲۷۱) and a researcher-made reading comprehension pretest (M = 24.12, SD = 4.11).

[Downloaded from mail.ijreeonline.com on 2026-11-18]

3.2 Instrument

3.2.1 English Language Proficiency Test

The OPT, version 1.1 UCLES (2001) was administered to check the homogeneity of the students.

3.2.2 Reading Comprehension Tests

The diagnostic reading comprehension tests, developed and validated by Azmoode et al.'s (2024) study, were used in this research. For developing the tests, the researchers benefited from procedures by Alderson et al. (1995) including: test specifications, item writing and revising, piloting and analysis, training the raters, monitoring raters' reliability, and validation. The test items were in "multiple-choice" and "short answer" format.

3.2.3 Reading Comprehension Passages

In the current study, the main reading materials were taken from well-known books including "Read This 2" (2010) and "For and against" (1968), the simplified version. The texts were selected in four different reading genres and they enjoyed the same difficulty level (11 on the average) which was computed through Fry's Readability Index.

3.2.4 Reading Comprehension Checklist

The reading comprehension assessment checklist developed by Azmoode et al. (2024) was used to assess the participants' rating accuracy, (see Appendix A). To develop the checklist, Bachman's (1990) guidelines for designing rating scales were used. The procedures were as follows:

- targeting the intended reading subskills
- defining the reading subskills operationally
- categorizing the abilities in five levels of performance
- describing the features of each performance level
- setting the cut-off score to each performance level

3.3 Procedure

The design of this study was quasi-experimental with two-group pretest-posttest. The OPT and reading pretest were administered at the outset of the study. Before that, the reading courses in the context of study had been explored. The course content and their requirements were examined thoroughly thus four widely- used reading genres in Iranian EFL tertiary education (argumentative, expository, descriptive, and narrative) were selected and a primary subskills list was created accordingly. To identify the learners 'challenging reading subskills, the students read some short texts, answered the comprehension questions and identified the questions that tested a particular subskill. The analysis of the students' responses revealed that distinguishing between "fact/opinion", "cause/effect", and "locating main ideas/supporting details" were the most difficult subskills for the learners.

As the next step, the students received self- and peer- assessment training and brief instruction to use the reading checklist for two sessions. In so doing, the instructor discussed the potential benefits of self- and peer- assessment and introduced the related techniques in each group. Next, the participants were provided with the reading comprehension assessment checklist based on which the components as well as the descriptors of each performance level were elaborated. In order to clearly establish the criteria of self- and peer-assessment, some graded samples were displayed on a video projector. The instructor clarified the process and the steps that had been taken to complete the sample checklists. The subjects then could use the samples as a model to rate their own, as well as their peers' reading comprehension. Being familiar with the checklist, the participants practiced the assessment of some short texts in both groups.

[Downloaded from mail.ijreeonline.com on 2026-11-18]

The instructional treatment took 12 weeks. The instructional procedures in this study encompassed three stages of pre-reading, while reading and post-reading. The pre-reading activities centered on the intervention of reading subskills. In fact, each pair of reading subskills was tested in four different reading genre texts, during four subsequent sessions (no genre instruction was intended in the design of the study). Next, the participants were provided with an unseen passage to preview. They were asked to read the text silently and answer the comprehension questions (while reading stage). Following that, the students in both groups filled out the reading checklist (post-reading stage). Every session, the instructor assessed the students' samples and provided them with feedback throughout the course; thus, the learners were informed about any evaluative mismatches between their own assessment and that of the instructor. Finally, the reading comprehension posttest was administered. In this study the reading comprehension tests required only objective judgments. To estimate the intra- rater consistency, 20 samples were randomly selected and the instructor rated them once again. A significant agreement between the first and second ratings (r (18) = .882) was shown.

3.4 Data Analysis

For analyzing the data descriptive statistics analysis was run and one-way between groups multivariate analysis of variance (MANOVA) was run to determine any statistical significance difference among self-, peer- and instructor-rating scores in the assessment of reading comprehension.

4. Results

Before testing the null hypotheses, the normality of the OPT was explored through skewness and kurtosis indices. The normality of the data was assured since the indices were lower than ±2 (George & Mallery, 2020). To address the research question, three sub-questions were formulated. For all data, the assumptions of equality of variance (by Levene's test) and homogeneity of variance-covariance matrices (by Box's test) were checked and none of them were violated.

4.1 Testing the First Subquestion

The first sub-null hypothesis was investigated through a test of MANOVA. The results show a statistically significant difference between the groups on the combined dependent variables, F (3, 60) = 6.24; Wilks' Lambda = .74, p = .001; partial eta squared = .25, which shows a large effect size and a substantial difference. The statistical significant difference, using a Bonferroni adjusted alpha level of .017, was in assessing main idea/ supporting idea subskills in argumentative genre (see Table 1). It indicates that the students in self-assessment group were only inaccurate in the assessment of main idea/ supporting idea subskill in the argumentative genre.

Table 1. Tests of between-subjects effects for self- and instructor- assessment scores in argumentative genre

Source	Dependent Variable	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Group	Comprehension1.Caus.Arg	1.838	1	1.838	1.867	.177	.031
	Comprehnsion5.Fact.op.Arg	3.384	1	3.384	3.625	.062	.059
	Comprhenson9.Main.sup.Arg	10.838	1	10.838	15.724	.000	.213

[Downloaded from mail.ijreeonline.com on 2026-11-18]

The result of MANOVA in descriptive genre shows a statistically significant difference between the groups on the combined dependent variables, F (3, 60) = 13.36; Wilks' Lambda = .58, p = .000; partial eta squared = .41, which indicates a large effect size showing a substantial difference. Considering the results of dependent variables separately, using a Bonferroni adjusted alpha level of .017, was in the assessment of all three subskills in descriptive genre. It reveals that the students in self-assessment group were inaccurate in the assessment of all the three subskills in the descriptive genre. In expository reading genre the results of MANOVA show a statistically significant difference between the groups on the combined dependent variables, F (3, 60) = 12.45; Wilks' Lambda = .60, p = .000; partial eta squared = .40, which shows a large effect size and a substantial difference. Bonferroni adjusted alpha level of .017 was in the assessment of cause/ effect, and main idea/supporting idea subskills in expository genre. This means that the learners in self-assessment group were inaccurate in performing the assessment of the cause/ effect, and main idea/ supporting idea subskills in the expository reading genre.

The results of MANOVA in narrative reading genre show a statistically significant difference between the groups on the combined dependent variables, F (3, 60) = 9.44; Wilks' Lambda = .66, p = .000; partial eta squared = .33, which shows a large effect size and a substantial difference. Bonferroni adjusted alpha level of .017 was in the assessment of cause/effect, and main idea/ supporting idea subskills in narrative genre (see Table 2). It suggests that the students in self- assessment group were inaccurate in performing the assessment of cause / effect, and main idea/ supporting idea subskills in the narrative genre.

Table 2. Tests of between-subjects effects for self- and instructor- assessment scores in narrative genre

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Group	Comprehension4.Caus.Nar	8.438	1	8.438	12.651	.001	.179
	Comprehnsion8.Fact.op.Nar	.551	1	.551	.683	.412	.012
	Comprhnson12.Main.sup.Nar	9.801	1	9.801	11.064	.002	.160

4.2 Testing the Second Sub-question

In argumentative genre the results show a statistically significant difference between the groups on the combined dependent variables, F (3, 60) = 10.95; Wilks' Lambda = .63, p = .000; partial eta squared = .37 which shows a large effect size and a substantial difference. Bonferroni adjusted alpha level of .017, was only in the assessment of main idea/ supporting idea subskill in argumentative genre (see Table 3). This means that the learners in peer- assessment group were inaccurate in performing the assessment of the main idea/ supporting idea subskill in the argumentative genre.

[Downloaded from mail.ijreeonline.com on 2026-11-18]

Table 3. Tests of between-subjects effects for peer- and instructor- assessment scores in argumentative genre

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Group	Comprehension 1.Caus.Arg	2.017	1	2.017	2.082	.154	.035
	Comprehnsion5.Fact.op.Arg	3.750	1	3.750	3.641	.061	.059
	Comprhenson 9.Main.sup.Arg	18.150	1	18.150	29.991	.000	.341

Another test of MANOVA was run in descriptive reading genre. The results show a statistically significant difference between the groups on the combined dependent variables, F (3, 60) = 10.95; Wilks' Lambda = .63, p = .000; partial eta squared = .37 which shows a large effect size and a substantial difference. Bonferroni adjusted alpha level of .017, was in the assessment of cause/ effect, and main idea/supporting idea subskills in descriptive genre. This means that the students in peer- assessment group were inaccurate in performing the assessment of cause/ effect and the main idea/ supporting idea subskills in the descriptive reading genre.

The results of MANOVA in expository genre show a statistically significant difference between the groups on the combined dependent variables, F (3, 60) = 17.26; Wilks' Lambda = .51, p = .000; partial eta squared = .48, which shows a large effect size and a substantial difference. Bonferroni adjusted alpha level of .017 was in the assessment of cause/effect, and main idea/supporting Idea subskills in expository genre (see Table 4). This means that the students in peer- assessment group were inaccurate in performing the assessment of the cause/ effect and the main idea/ supporting idea subskills in the expository reading genre.

Table 4. Tests of between-subjects effects for peer- and instructor- assessment scores in expository genre

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Group	Comprehension3.Cause.Ex	21.004	1	21.004	27.351	.000	.320
	Comprhension7.Fact.op.Ex	4.134	1	4.134	5.765	.020	.090
	Compheson11.Main.sup.Ex	12.604	1	12.604	16.764	.000	.224

In narrative genre, the results show a statistically significant difference between the groups on the combined dependent variables, F (3, 60) = 8.65; Wilks' Lambda = .68, p = .000; partial eta squared = .31, which shows a large effect size and a substantial difference. Bonferroni adjusted alpha level of .017, was in the assessment of cause/ effect, and main idea/ supporting idea subskills in narrative genre. This means that the students in peer- assessment group were inaccurate in performing the assessment of cause/ effect and the main idea- supporting idea subskills in the narrative reading genre.

[Downloaded from mail.ijreeonline.com on 2026-11-18]

4.3 Testing the Third Sub-question

In order to make comparisons between self- and peer- assessment scores in assessing the reading subskills, the same procedures were used. In argumentative reading genre the results show no statistically significant difference between the groups on the combined dependent variables, F (3, 60) = 1.16; Wilks' Lambda = .94, p = .333; partial eta squared = .05, which shows a weak effect size and a small difference. This means that there is no significant difference between the two groups in the assessment of the subskills in the argumentative reading genre. The results of MANOVA in descriptive reading genre show no statistically significant difference between the groups, F (3, 60) = .18; Wilks' Lambda = .99, p = .904; partial eta squared = .01, which shows a weak effect size and a small difference. This means that there is no statistically significant difference between the two groups in the assessment of the subskills in the descriptive reading genre. In expository genre the results show no statistically significant difference between the groups, F (3, 60) = .18; Wilks' Lambda = .99, p = .904; partial eta squared = .01, which suggests a weak effect size indicating a small difference. It means that there is no statistically significant difference between the two groups in the assessment of three subskills in the expository reading genre. In narrative genre the results of MANOVA show no statistically significant difference between the groups, F (3, 60) = .18; Wilks' Lambda = .99, p = .904; partial eta squared = .01, which shows a weak effect size and a small difference. It means that there is no significant difference between the two groups in the assessment of three subskills in the narrative reading genre. Thus, the null hypothesis was rejected suggesting that there was statically significant difference between the accuracy of self-, peer- and instructor -ratings in assessing reading comprehension of learners in four reading genres.

5. Discussion

The present study was conducted to examine EFL learners' rating accuracy in assessing reading comprehension in various genres which ultimately resulted in some diagnostic information. While the results show the inaccuracy of self- and peer- assessment, the evaluative mismatches between the students- and those of the instructors' assessment throughout the course raised the learners' awareness regarding their strengths and weaknesses in reading comprehension subskills and helped them overcome their weaknesses. More importantly, as it was confirmed previously, engaging learners in assessment process can help them develop a critical view in learning and foster their autonomy in language learning over time (Butler, 2018; Paris & Paris, 2001).

According to the findings, the two groups were inaccurate in assessing the subskill of main idea/supporting details in all genres and the subskill of cause / effect was the second most difficult subskill for them. This result is quite compatible with past research, in the domain of cognitive diagnostic assessment, which confirmed that the subskills of cause/effect and main idea/supporting details were the least-mastered subskills by learners (Javidanmehr & Anani Sarab, 2019; Ravand, 2015). Based on the findings, while the peer- assessment group was never inaccurate in assessing fact/opinion subskill, the self-assessment group was inaccurate in assessing fact and opinion only in the descriptive genre; it can be concluded that this subskill was easier for the learners to assess. In fact, conducting detailed analysis and providing such diagnostic information can help both learners and instructors identify the problematic areas in students' performance in order to target them in instruction and elevate the students' weaknesses.

The learners' inaccuracy in reading comprehension assessment can also be associated with the challenging nature of the reading genres since in the present study the learners had assessment inaccuracy in almost all genres. Contrary to this finding, some other studies (Carrell & Connor, 1991; DuBravac & Dalle, 2002; Sahin, 2013) proved that narrative and descriptive texts were easy to comprehend for learners. Rouhi et al. (2015) in their study confirmed that argumentative and expository reading genres were more challenging than descriptive and narrative genres for EFL learners. Although the reading passages in this study enjoyed the same level of difficulty, it has been suggested that the variations in text genres might lead to the learners' trouble in reading comprehension (Zhou & Siriyothin, 2011);

[Downloaded from mail.ijreeonline.com on 2026-11-18]

this was previously confirmed by investigating the effects of text types (expository and narrative) on improving students' reading comprehension. Since in Zhou and Siriyothin's study (2011) the students had a better performance on expository than narrative texts, it was inferred that types of genre had remarkable effects on learners' reading comprehension. Further, the relative effects of various genres and subskills in reading tests were examined previously and it was proved that the inclusion of both reading genres and subskills in instruction could have a positive effect on the learners' reading scores reliability (Shin, 2002).

Regarding the subskills approach, the present research is in line with past empirical findings in which reading comprehension is viewed as a divisible construct, containing several subskills or micro-skills (Karakoc, 2019; Kim & Jang, 2009; Pan, 2009; Shin, 2002). Taking the subskills approach, this study supports the pan's findings (2009) which suggest that through subskills approach learners are equipped with essential skills for better reading comprehension and by conscious practice students can find mastery in certain reading skills.

6. Conclusion

Taking a subskills approach and involving the students in the assessment of reading comprehension, this study revealed areas in the learners' performance that called for improvement. The inaccuracy of the students in assessments of reading comprehension pinpoints the fact that not only the learners need more training and practice in applying self-, and peer- assessment techniques, but also they need to know more about specific reading subskills. Furthermore, despite their assessment inaccuracy, the students in this study gained awareness regarding their own difficulties in certain reading comprehension subskills and genres by receiving the related diagnostic feedback.

The present study bears both theoretical and practical implications for EFL instruction. This study can add more weights to the theoretical foundation of the diagnostic assessment theory (Alderson et al., 2015) by confirming that, engaging the learners in assessment can substantially help them find awareness regarding their own strengths and weaknesses in reading comprehension. The present research has some practical implications for the main EFL stakeholders such as test developers and materials designers. Test developers can turn diagnostic information into some applicable standards to contribute to modifying the teaching and learning pedagogy (Javidanmeh & Anani Sarab, 2019). Material developers can also target the most challenging subskills and genres in course books; so that they can function as agents of change.

The current research suffers from some limitations; first, the findings of this study would be more generalizable if they were confirmed in replication, focusing on various subskill types in different reading genres. Second, the variable of students' proficiency level was not considered in this research; therefore, the effect of diverse proficiency levels on the accuracy of learners' self- and peer- assessments requires more research attention. For future exploration, research on various reading subskills and genres is recommended. More importantly, the actual application of diagnostic information in EFL instruction is an area that warrants more research (Liu, 2014).

[Downloaded from mail.ijreeonline.com on 2026-11-18]

References

Alderson, J. C., Clapham, C. M., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge University Press.

Alderson, J. C., Brunfaut, T., & Harding, L. (2015). Towards a theory of diagnosis in second and foreign language assessment: Insights from professional practice across diverse fields. *Applied Linguistics, 36*(2), 236- 260. https://doi.org/10.1093/applin/amt046

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. https://doi.org/10.1016/j.system.2004.04.006

Alibakhsh, G. (2013). Construction and validation of self-assessment inventory for English for academic purposes: A case of Iranian tertiary students. *RALs, 4*(2), 93-109.

Aminu, N., Hamdan, M., & Russell, C. H. (2021). Accuracy of self-evaluation in a peer- learning environment: an analysis of a group learning model. *SN Social Science, 1*(185), 1- 17. https://doi.org/10.1007/s43545-021-00152-3

Aryadoust, V. (2019). A review of comprehension subskills: A scientometrics perspective. *System, 88*, 1-16. https://doi.org/10.1016/j.system.2019.102180

Ashton, K. (2014). Using self-assessment to compare learners' reading proficiency in a multilingual assessment framework, *System, 42*, 105–119. https://doi.org/10.1016/j.system.2013.11.006

Azmoode, M., Kiany, G. R., & Abbasian, G. R. (2024a). Diagnostic assessment of interactional competence in paired speaking tests: Investigating rating accuracy of Iranian EFL learners. *Journal of Language and Translation, 4*(1), 19-33. https://doi.org/10.30495/ttlt.2024.709364

Azmoode, S. M., Kiany, G. R., & Abassian, G. R. (2024). On the effect of diagnostic self-, and peer- assessment on reading comprehension: Examining EFL learners' diagnostic rating accuracy across various genres. *Journal of Modern Research in English Language Studies, 11*(2), 177-202. https://doi.org/10.30479/jmrels.2023.18703.2204

Bachman, L. F. (4002). *Fundamental considerations in language testing*. Oxford University Press.

Bailey, A. L., & Heritage, M. (2018). *Self-Regulation in learning: The role of language and formative assessment*. Harvard Education Press.

Butler, Y. G. (2018). The role of context in young learners' processes for responding to self- assessment items. *The Modern Language Journal, 102*(1), 1-20. https://doi.org/10.1111/modl.12459

Benson, P., & Voller, P. (2014). *Autonomy and independence in language learning*. Routledge.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment Evaluation and Accountability, 21*(1), 5-31. doi:10.1007/s11092-008-9068-5

Butler, Y. G., & Lee, J. (2006). On-task versus off-task self-assessments among Korean elementary school students studying English. *The Modern Language Journal*, *90*(4), 506–518. https://doi: 10.1111/j.1540–4781.2006.00463.x

Butler, Y. G., & Lee, J. (2010). The effects of self-assessment among young learners of English. *Language Testing, 27*(1), 5–31. https://doi.org/10.1177/0265532209346370

Carrell, P. L., & Connor, U. (1991). Reading and writing descriptive and persuasive texts. *Modern Language Journal, 75*(3), 314-324. https://doi.org/10.1111/j.1540-4781.1991.tb05361.x

[Downloaded from mail.ijreeonline.com on 2026-11-18]

Chen, Y. M. (2008). Learning to self-assess oral performance in English: longitudinal case study. *Language Teaching Research*, *12*(2), 235-262. https://doi.org/10.1177/1362168807086293

Davis, F. B. (1944). Fundamental factors of comprehension in reading. *Psychometrika, 9*(3),185-197. https://doi.org/10.1007/BF02288722

Dörnyei, Z., & Ushioda, E. (2013). *Teaching and researching motivation*. Routledge.

DuBravac, S., & Dalle, M. (2002). Reader question formation as a tool for measuring

comprehension: Narrative and expository textual inferences in a second language. *Journal of Research in Reading, 25*(2), 217-231. https://doi.org/10.1111/1467-9817.00170

Earl, L. M. (2012). *Assessment as learning: Using classroom assessment to maximize student learning*. Corwin Press.

Elahi, M. S. (2016). Assessing and improving general English university students' main sub-skills of reading compression: A case of university of Bojnord. *Sino-US English Teaching, 13*(4), 245-260. https://doi:10.17265/1539-8072/2016.04.002

Esfandiari, R., & Jafari, H. (2021). Morphological complexity across descriptive expository, and narrative text types in Iranian lower- intermediate language learners. *Issue in Language Teaching, 10*(1), 237-267. https://doi.org/10.22054/ilt.2021.59736.580

Esfahani, K. M., Rashtchi, M., Abousaidi, R. A., & Mowlaie, B. (2022). Promoting metacognitive awareness in writing assessment tasks through planning, monitoring, and evaluation: Achievements and perceptions. *Iranian Journal of English for Academic Purposes, 11*(3), 50-68. dor: 20.1001.1.24763187.2022.11.3.4.9

Farhady, H., & Daftarifard, P. (2006). On the scalability of the components of the reading comprehension ability: A progress report. In H. Farhady (Ed.), *Twenty-five years of living with applied linguistics: Collection of articles* (pp. 189-204). Rahnama publisher.

George, D., & Mallery, P. (2020). *IBM SPSS statistics 26 step by step: A simple guide and reference*. Routledge.

Goodman, K. S. (1967). Reading: A psycholinguistic guessing game. *Journal of the Reading Specialist, 6*(4), 126-135. https://doi.org/10.1080/19388076709556976

Goral, D. P., & Bailey, A. L. (2019). Student self-assessment of oral explanations: Use of language learning progressions. *Language Testing, 36*(3), 1- 27. https://doi.org/10.1177/0265532219826330

Green, A. (2018). Assessment for learning in language education**.** *Iranian Journal of Language Teaching Research, 6*(3), 9-18. https://files.eric.ed.gov/fulltext/EJ1192609.pdf

Han, C., & Riazi, M. (2018). The accuracy of student self-assessments of English-Chinese bidirectional interpretation: A longitudinal quantitative study. *Assessment & Evaluation in Higher Education*, *43*(3), 386-398. https://doi.org/10.1080/02602938.2017.1353062

Jang, E. E. (2009). Demystifying a Q-Matrix for making diagnostic inferences about L2 reading skills. *Language Assessment Quarterly, 6*(3), 210-238. https://doi.org/ 10.1080/ 15434300903071817

Javidanmehr, Z., & Anani Sarab, M. R. (2019). Retrofitting non- diagnostic reading comprehension assessment: Application of the G- DINA model to a high stakes reading comprehension test. *Language Assessment Quarterly, 16(3),* 294-311. https://doi.org/ 10.1080/15434303.2019.1654479

Karakoc, A. I. (2019). Reading and listening comprehension Subskills: The match between theory, course books, and language proficiency tests. *Advances in Language and Literary Studies (ALLS), 10*(4), 166_171. https://doi.org/10.7575/aiac.alls.v.10n.4p.166

[Downloaded from mail.ijreeonline.com on 2026-11-18]

Khalifa, H., & Weir, C. J. (2009). *Examining reading: Research and practice in assessing second language reading*. Cambridge University Press.

Kim, A. (2015). Exploring ways to provide diagnostic feedback with an ESL placement test: Cognitive diagnostic assessment of L2 reading ability. *Language Testing, 32*(2), 227-258. https://doi.org/10.1177/0265532214558457

Kim, Y. H., & Jang, E. E. (2009). Differential functioning of reading subskills on the OSSLT for L1 and ELL students: A multidimensionality model-based DBF/DIF approach. *Language Learning, 59*(4), 825–865. doi:10.1111/j.1467-9922.2009.00527.x

Lave, J. (1993). The practice of learning. In S. Chaiklin & J. Lave (Eds.), *Understanding practice. Perspectives on activity and context.* Cambridge University Press.

Lee, Y. W. (2015). Diagnosing diagnostic language assessment. *Language Testing*, *32*(3) 299 –316. doi:10.1177/0265532214565387

Lee, S. K., & Chang, S. H. (2005). Lerner involvement in self- and peer- assessment of task- based oral performance. *Second Language Research*, *41*, 711-735. file:///C:/Users/SMA/Downloads/10.+2231654.pdf

Lee, Y. W., & Sawaki, Y. (2009). Application of three cognitive diagnosis models to ESL reading and listening assessments. *Language Assessment Quarterly, 6*(3), 239-263. https://doi.org/10.1080/15434300903079562

Liu, H. H. (2014). The conceptualization and operationalization of diagnostic testing in second and foreign language assessment. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics, 14*(1), 1-12.

Liu, H. H., Alderson, J. C., & Brunfaut, T. (2015). Diagnostic assessment of reading and listening in a second or foreign language: Elaborating on diagnostic principles. *Language Testing, 32*(3), 317-336. https://doi.org/10.1177/0265532214564505

Lu, L. (2018). An analysis of peer-assessment in Chinese as a second language classroom presentation. *Chinese Language Teaching Methodology and Technology*, *1*(3), 18. https://engagedscholarship.csuohio.edu/cltmt/vol1/iss3/3

Mazloomi, S., & Khabiri, M. (2016). Diagnostic assessment of writing through dynamic self-assessment. *International Journal of English Linguistics*, *6*(6), 19-31. doi:10.5539/ijel.v6n6p19

Ma, W., & Winke, P. (2019). Self-assessment: How reliable is it in assessing oral proficiency over time? *Foreign Language Annals, 52(1),* 66-86. https://doi.org/10.1111/flan.12379

Markey, M. (2020). Using diagnostic assessment to investigate challenges in second language reading. *The Reading Matrix: An International Online Journal, 20*(1), 45-65. https://readingmatrix.com/files/22-045p942v.pdf

Nalbantoğlu, Y, F. (2017). Reliability of scores obtained from self-, peer-, and teacher-assessments on teaching materials prepared by teacher candidates. *Educational Sciences: Theory & Practice*, *17*, 395–409. http://dx.doi.org/10.12738/ estp.2017.2.0098

Ng, M. C. W. (2018). Assessment for/as learning in Hong Kong English language classrooms: A review. *IJREE*, *3*(3), 1-12. doi:10.29252/ijree.3.3.1 http://ijreeonline.com/article-1-104-en.html

Oscarson, M. (2013). Self-assessment in the classroom. In A. J. Kunnan (Ed), *The companion to language assessment* (pp. 712-729). Wiley-Blackwell

[Downloaded from mail.ijreeonline.com on 2026-11-18]

Ou, W. J. A. (2022). Writing accessible theory in ecology and evolution: Insights from cognitive load theory. *BioScience*, *72*(3), 300–313. https://doi.org/10.1093/biosci/biab133

Oxford, R. L. (2016). *Teaching and researching language learning strategies: Self-regulation in context*. Routledge.

Paleczek, L., Seifert, S., Schwab, S., & Gasteiger-Klicpera, B. (2015). Assessing reading and spelling abilities from three different angles–correlations between test scores, teachers' assessment and children's self-assessments in L1 and L2 children. *Procedia - Social and Behavioral Sciences, 174*, 2200-2210. https://doi.org/ 10.1016/j.sbspro.2015.01.876

Pang, N. S. (2020). Teachers' reflective practices in implementing assessment for learning skills in classroom teaching. *ECNU Review of Education*, 1–21. doi:10.1177/2096531120936290

Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self- regulated learning. *Educational Psychology, 36*(2), 89–101. https://doi.org/ 10.1207/S15326985EP3602_4

Ravand, H. (2015). Application of a cognitive diagnostic model to a high-stakes reading comprehension test. *Journal of Psychoeducational Assessment, 34*(8), 782–799. https://doi.org/ 10.1177/0734282915623053

Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research, and Evaluation, 11*(10), 1-13. doi: https://doi.org/10.7275/9wph-vv65

Rost, D. H. (1993). Assessing different components of reading comprehension: Fact or fiction? *Language Testing, 10*(1), 79–92. https://doi.org/ 10.1177/026553229301000105

Rouhi, A., Jafarigohar, M., Alavi, M., & Hosseini. Y. (2015). Task difficulty of macro-genres and reading strategies and reading comprehension. *International Journal of Asian Social Science, 5*(11), 656-677. https://doi.org/ 10.18488/journal.1/2015.5.11/1.11.656.677

Şahin, A. (2013). The effect of text types on reading comprehension. *Mevlana International Journal of Education, 3*(2), 57-67. https://doi.org/ 10.13054/mije.13.27.3.2

Santos, L., & Semana, S. (2015). Developing mathematics written communication through expository writing supported by assessment strategies. *Educational Studies in Mathematics*, *88*(1), 65-87. http://doi.org/10.1007/s10649-014-9557-z

Sawaki, Y., Kim, H. J., & Gentile, C. (2009).Q-matrix construction: Defining the link between constructs and test items in large-scale reading and listening comprehension assessments. *Language Assessment Quarterly, 6*(3), 190– 209. https://doi.org/10.1080/15434300902801917

Schunk, D. H. (2004). *Learning theories: An educational perspective*. Merrill Prentice.

Singh Negi, J., & Laudari, S. (2022). Challenges of developing learner autonomy of english as a foreign language (EFL) learners in underprivileged areas. *IJREE*, *7*(2), 65-80. doi:10.52547/ijree.7.2.65 http://ijreeonline.com/article-1-689-en.html

Song, M. (2008). Do divisible subskills exist in second language (L2) comprehension? A structural equation modeling approach. *Language Testing, 25*(4), 435–464. https://doi.org/10.1177/0265532208094272

Spearritt, D. (1972). Identification of subskills of reading comprehension by maximum likelihood factor analysis. *Reading Research Quarterly, 8*(1), 92-111.

Surbakti, R., Umboh, E. S., Pong, M., & Dara, S. (2024). Cognitive load theory: Implications for instructional design in digital classrooms. *International Journal of Educational Narrative*, *2*(6), 483-493. https://doi.org/10.70177/ijen.v2i6.1659

[Downloaded from mail.ijreeonline.com on 2026-11-18]

Suzuki, Y. (2015). Self-assessment of Japanese as a second language: The role of experiences in the naturalistic acquisition. *Language Testing, 32*(1) 63–81. doi:10.1177/0265532214541885

Tengberg, M. (2018). Validation of sub-constructs in reading comprehension tests using teachers' classification of cognitive targets, *Language Assessment Quarterly*, 1-13. https://doi.org/10.1080/15434303.2018.1448820

Toledo, P. F. (2005). Genre analysis and reading of English as a foreign language: Genre schemata beyond text typologies. *Journal of Pragmatics, 37*, 1059–1079. https://doi.org/ 10.1016/j.pragma.2005.01.002

Vacca, R. T. (1980). A study of holistic and subskill instructional approaches to reading comprehension. *Journal of Reading, 23*(6), 512–518. https://www.jstor.org/stable/40028836

Vygotsky, L. (1978). *Mind in society. The development of higher psychological processes*. Harvard University Press.

van Kraayenoord, C. E., & Paris, S. G. (1997). Australian students' self-appraisal of their work samples and academic progress. *The Elementary School Journal*, *97*(5), 523–537. https://doi.org/10.1086/461879

Xiao,Y., & Lucking, R. (2008). The impact of two types of peer assessment on students' performance and satisfaction within a Wiki environment. *Internet and Higher Education, 11,* 186-193.

Yin, J. (2018). A review on researches of "genre-based teaching approaches" in recent 20 years in China. *International Journal of Secondary Education, 6*(1), 16-23. https://doi.org/ 10.11648/j.ijsedu.20180601.14

Yoshida, M. (2012).The interplay of processing task, text type, and proficiency in L2 reading *Reading in a Foreign Language, 24*(1), 1-29. https://files.eric.ed.gov/fulltext/EJ974102.pdf

Zhou, L., & Siriyothin, P. (2011). Effects of text types on advanced EFL learners' reading comprehension. *Journal of Language and Culture, 30*(2), 45-66.

[Downloaded from mail.ijreeonline.com on 2026-11-18]