

## Perceptual (Mis)Matches between Learners' and Teachers' Rating Criteria in the Iranian EFL Writing Self-assessment Context

Razieh Mohammadi<sup>1</sup>, Nasim Ghanbari<sup>2\*</sup>, Abbas Abbasi<sup>3</sup>

ARTICLE INFO	ABSTRACT
<p><b>Article History:</b> Received: August 2023 Accepted: September 2023</p> <hr/> <p><b>KEYWORDS</b> Perceptual mismatch Rating criteria Self-assessment Teacher assessment Writing assessment</p>	<p><b>Abstract</b> As a formative assessment procedure, self-assessment aims to converge learners' and teachers' views in assessment. Hence, reducing the perceptual mismatches between the learners' and the teachers' assessments would positively affect the learning process. For this aim, the present study investigated to what extent the learners' assessment of their writing before and after being provided with a list of rating criteria, agrees with that of their teachers. Therefore, a body of six EFL writing teachers and 27 EFL learners participated in this study. The learners were asked to rate their writing before and after receiving the rating criteria developed by the researchers. The teachers also rated the students' writings following the same criteria. The obtained results showed a significant difference between the students' scores on the first and second assessment occasions. The teachers' and the students' assessments on the second time also were found to significantly correlate. Moreover, the analysis of the students' comments showed that while they rated their writing on some limited aspects of writing in the first rating occasion, they assessed their essays using more components in the second assessment phase. Overall, the findings revealed that providing the learners with rating criteria would not only reduce the perceptual mismatches between the students' and the teachers' assessment but promote a more democratic classroom assessment. The findings of the study reduce the complexity of self-assessment practice by narrowing the perceptual gap between the students and the teachers.</p>

### 1. Introduction

Improving the standards of learning is strongly pursued in many educational curricula across the world. Learning is the outcome of what teachers and students do in the classroom; therefore, instructors and students need to see learning as an interactive process (Arnold, 2022). Adopting this view, the teachers

<sup>1</sup> Department of English language and literature, Faculty of Humanities, Persian Gulf University, Bushehr 75169, Iran, Email: [raziyemhmdi21@gmail.com](mailto:raziyemhmdi21@gmail.com)

<sup>2</sup> Department of English language and literature, Faculty of Humanities, Persian Gulf University, Bushehr 75169, Iran, Email: [btghanbari@pgu.ac.ir](mailto:btghanbari@pgu.ac.ir)

<sup>3</sup> Department of English language and literature, Faculty of Humanities, Persian Gulf University, Bushehr 75169, Iran, Email: [abbasi42@gmail.com](mailto:abbasi42@gmail.com)

Cite this paper as: Mohammadi, R., Ghanbari, N., & Abbasi, A. (2024). Perceptual (mis)matches between learners' and teachers' rating criteria in the Iranian EFL writing self-assessment context. *International Journal of Language Testing*, 14(1), 150–165. <https://doi.org/10.22034/IJLT.2023.410811.1274>

constantly make an effort to monitor students' needs and adapt their practices in designing and improving the related curriculum (Kumaravadivelu, 2003; Orsmond, 2011). Johnson (1989, p.1) broadly defines a curriculum as one which "includes all the relevant decision-making processes of all the participants." He believes that a coherent curriculum is the one in which all the participants including material writers, teacher trainers, curriculum designers, policymakers, teachers, and learners have a convenient understanding of the objectives and goals for which the curriculum is designed. In other words, in a coherent curriculum, the attempt is made to constantly reduce the perceptual mismatches between the teacher and the learners.

Hence, coherence is set to become a vital factor in any successful curriculum. Moreover, identifying mismatches can help teachers and students. It also aids the success of particular curricula (Babaii et al., 2016; Kumaravadivelu, 2003).

In the words of Kumaravadivelu (1991), the mismatches probably result from what he calls teacher intention and learner interpretation. Needless to say, the narrower the gap between these two, the more success will result in achieving curriculum goals in teaching. Therefore, in order to maintain a coherent curriculum, teachers' and students' responsibility is to create and promote learning opportunities through meaningful interaction, negotiation, and collaboration (Johnson, 1989; Kumaravadivelu, 1991).

Self-assessment, as one kind of formative assessment practice, provides a context in which students are active and responsive for their own learning (Arnold, 2022). Moreover, they can observe the relationship between the meaning and the task (Bourke & Mandia, 2013). According to Luoma and Tarnanen (2003), alternative aspects of learning-centered assessment merge the process of learning and assessment; thus, it can raise students' awareness of their learning objectives through feedback and assessment. Hence, self-assessment can be considered a robust alternative to the traditional way of evaluation, which is always teacher-fronted. It promotes the students' agency by emphasizing their voice and choice in the assessment tasks. As a result, many studies recommend using self-assessment along with teacher assessment in classroom evaluation tasks (Boud, 2013; Bourke & Mentis, 2013; Falchikov, 2013).

The existing literature shows that few studies have investigated the mismatches between the teachers' and students' evaluation criteria. Among the few studies and in the context of speaking self-assessment, Babaii et al. (2016) investigated the learners' rating criteria for L2 speaking and compared them with those of their teachers. In another study, Tajgozari and Alimorad (2019) explored the effect of (mis)mismatches between EFL teachers' and students' perceptions of L2 writing assessment on students' writing achievement scores. Moreover, the majority of the studies have investigated teachers' and students' perceptions of the academic writing task (Wolsey et al., 2012). Other studies also have focused on the learners' perceptions of written corrective feedback in writing assessment (Eun, 2015; Montgomery & Baker, 2007).

Despite the benefits counted for the self-rating in writing assessment, there is still an assumption that teacher scoring is always correct; whereas this may not always hold true. In fact, in a test made up of items for which there is only one correct answer, the traditional approach could be appropriate to be applied; however, in performance tests, such as written compositions, the use of a single assessor may result in a biased evaluation (Matsuno, 2009). Moreover, since the ability to write effectively has received increasingly global importance, the instructor role has proven to be essential in both second and foreign language learning (Weigle, 2002). Weigle states that if the teachers consider the emphasized valuable aspects of writing in criteria, it will enable students to work on these aspects of writing and also, they will be aware of the way their works will be scored. The quality of writing assessment depends upon the criteria used in assessing a piece of writing. For this reason, using a scoring procedure allows for more discussion about the objectives of writing and expected outcomes for the students.

However, since the students can have a different interpretation of some criteria in comparison to their teachers, the students' involvement in the activity of assessing their own performance may still raise the mismatches in the classroom. According to Orsmond, et al. (2000), a factor that is helpful to solve these problems and to achieve the link between tutor assessment and student learning is to establish a common understanding of the assessment criteria between the tutor and the learner. Therefore, this study aimed to investigate any differences between learners' assessment before and after their being provided with a list

of rating criteria and a practice session. Moreover, the study examined the extent to which the administration of the scoring criteria affected the congruence between the learners' and the teachers' writing assessments.

## 2. Review of Literature

### 2.1. *Perceptual Mismatch*

In the words of Brown (2006), perceptions refer to subjective beliefs and evaluations which play a crucial role in teaching and learning as they form the basis of teachers' and students' beliefs, opinions, and actions. Consequently, differences in teachers' and students' perceptions would negatively affect the outcome of learning and evaluation. In the same line, Barkhuizen (1998) believes that learner perception involves learners in a decision-making process. Writing assessment as a particular kind of performance assessment proceeds in an interactive way by involving both students and teachers. In this interactive decision-making cycle, teachers continuously make global and local decisions about the students' performance and the criteria they use in their assessment. Students also have a prominent role in this process. In fact, the way they perceive the assessment construct and how the teachers conduct their assessment considerably determine the success of the assessment practice.

According to Kumaravadivelu (1991), there are ten potential sources of mismatches between teacher intention and learner interpretation. These include cognitive mismatch (general cognitive knowledge of the world), communicative mismatch (communicative skills needed to convey messages), linguistic mismatch (syntactic, semantic, and pragmatic knowledge of the target language), pedagogic (teacher and learner perceptions of stated or unstated short- or long-term instructional objective(s) of language learning tasks (p.83), strategic mismatch ( learning strategies and styles), cultural mismatch (prior knowledge of the cultural norms of the target language community), evaluative mismatch ( articulated or unarticulated types of self-evaluation measures used by learners to monitor their ongoing progress in their language-learning activities (p. 86), procedural mismatch (stated or unstated paths chosen by learners to do a task), instructional mismatch (instructional guidance given by the teacher or by the textbook writer for successful completion of the task) and attitudinal mismatch ( participants' attitudes toward the nature of L2 learning and teaching, the nature of classroom culture, and teacher- learner role relationships).

Considering the fact that "perceptual mismatches are unavoidable, identifiable and manageable" (Kumaravadivelu, 2003, p.90), the first step is the identification of different kinds of mismatches that frequently occur in a language class through careful observation and meticulous analysis of teacher-learner interactions. The next step will be adopting context-specific techniques and strategies by teachers to reduce the degree of gap that exists between their intention and their learners' interpretation.

Ramazani (2014) who aimed to examine the characteristics of an effective teacher from the views of teachers and students found that there was a great mismatch between teachers' and students' perceptions. While teachers considered English proficiency as the most important criterion, students perceived pedagogical knowledge as a significant factor.

In another study, Lee and Choi (2019) investigated mismatches between students and teachers in Korean secondary schools. The results collected through questionnaires and interviews indicated areas of similarity and mismatches between the students and the teachers. While both students and teachers have similar perceptions of less frequent involvement in the communicative approach, the teachers perceived that non-communicative teaching practices occurred more frequently than did the students.

Harper and Puji Widodo (2020) also investigated the perceptual mismatches in the interpretation of task-based ELT materials in the context of a Chinese university. The qualitative analysis of a task-based English lesson showed that there was an agreement in the students' perception that the lesson focused more on meaning and less on form. The study concluded that learning goals should be considered as a shared vision and students should have a role in the design of classroom tasks.

Sengsouliya et al. (2021) explored the mismatches in teachers' teaching styles and students' learning style preferences. The study found areas of mismatches in styles between teachers' teaching styles and students' learning style preferences in four out of the five schools under study.

The existing research on writing self-assessment shows that when rating scales are shared with the student writers to assess their writing, they actively take responsibility for the assessment task (Ghanbari,

2019). Despite the importance of the possible evaluative mismatches in learners' and teachers' perceptions of writing ability, to the best of our knowledge, no study has ever investigated the evaluative gaps in writing self-assessment. Hence, the present study was an attempt to reveal the inside of learner-teacher interaction in the writing self-assessment practice.

## **2.2. Writing Self-assessment**

Teachers and language instructors are responsible for providing the students with appropriate instruction to meet their learning needs. To achieve this goal, an integrated context which reduces the mismatches in classrooms by using interactive practices is required. One way could be the focus on alternative assessment methods which emphasize the replacement of traditional approaches with more collaborative, democratic, and formative assessment practices (Manzari, 2023). Self-assessment is considered as one of the alternative assessment procedures. According to Boud (1995, p.5), the defining characteristic of self-assessment is "the involvement of the students in identifying standards and/or criteria to apply to their work and making judgment about the extent to which they have met these criteria and standards." Previous research shows that pupils can be highly reliable and honest in assessing themselves. They are also strict on themselves in assessing their own work (Black & Williams, 1998). It is said to be beneficial because everyone is trying to learn feedback about their desired goals and their current position and also the ways to close the gap between these two (Sadler, 1989).

Self-assessment has received increasing attention since not only does it provide an opportunity for learners to take an active role in understanding their strengths and weaknesses but also it may help the instructors to convey the intended messages about what students need to do in order to fill the gap between their present position and their future standards, as well as reducing the mismatches through involving the students in the assessment tasks.

Mutual understanding of classroom objectives between learners and teachers is of great importance as it leads to a coherent and ultimately more successful curriculum (Johnson, 1989; Kumaravadivelu, 1991). The collaboration and negotiation of teachers and students in the classroom provide students with an awareness of the objectives of the curriculum as well as the process of learning; hence they will be able to have an active role in their own learning process. Using criteria that are negotiated between the teacher and students decreases the probable mismatches in teachers' intentions and learners' interpretation of curriculum objectives in the evaluation. The reduction of these inconsistencies is believed to be a probable solution for a successful curriculum.

Nicol and Macfarlane-Dick (2006) emphasized the importance of teacher and learner discussions about the assessment process and the rating criteria which they believed could enhance the quality of assessment and reduce the evaluative gap between teacher and learners in the assessment tasks. In this regard, Orsmond (2011) introduced the GOALS process to underscore the role of learning in self-assessment. In the GOALS process, learners make concerted efforts to achieve the learning objectives by using self-regulation strategies and constantly evaluating their learning.

Despite the benefits associated with self-assessment, its role is still under-represented in many educational contexts (Williams, 1992). This is primarily because self-assessment is a hard goal to achieve, as it requires the control of both instrument and outcome by the teacher (Somervell, 1993). In order to control the instrument, teacher and learner should come to a common understanding of the standards of assessment as making criteria is a key success to any assessment procedure. Scholars also claimed that a factor that is helpful in achieving the link between teacher assessment and student learning is to provide a common understanding of the assessment criteria between the teacher and learners (Orsmond et al., 2000).

Scoring rubric is a critical component of assessment. This is because it represents the definition of the construct that the assessment intends to measure (Weigle, 2002). Despite these claims about a common understanding of assessment standards, the number of studies that investigated the mismatches between student and teacher performance in rating is extremely limited. The case is that getting a better understanding demands empirical research to examine the differences in learner and teacher perception of the assessment criteria. However, most studies conducted were to investigate the differences between professional raters and novice ones as well as the comparison of native and non-native raters. Although

Matsuno (2009) conducted a quantitative study on self- and peer-assessment, in comparison to teachers, there is an urgent need for further studies that focus on the differences between learners as raters with the tutor as a rater in writing, which is still underrepresented in the EFL context of writing classrooms. Many studies suggest that for students to gain a better understanding of rating goals, they should have an active role in constructing criteria (Buhagiar, 2007; Stefani, 1994).

Li and Lindsey (2015) showed how discrepancies between students' and teachers' interpretations of rubrics might influence the use of rubrics. The results revealed while the instructors focused convergently on a core set of focal points in the performance descriptions of the rubric, the students encountered difficulties in recognizing and/or explaining the focal points. Besides, in comparison to instructors', students' scores varied to a greater extent because they claimed that they were confused about the rubric. Their findings showed that classroom time particularly devoted to discussing and understanding rubrics is needed.

In the same line, Becker (2016) found that those students who participated in the development and/or application of the scoring rubric had significantly higher scores on the post-test summary writing task. The study argued that co-constructing, discussing and even using a rubric lead to more critical thinking about one's writing performance.

Ghanbari (2019) shared the rating scale with the students in an Iranian EFL writing assessment context. The study aimed to find out if sharing the rating scales with the students would reduce the perceptual mismatches of evaluation criteria in their writing. The results of the study showed noticeable progress in the performance of the students who had received the rating scale. In the same vein, Manzano (2022) examined the interrater reliability between self- and teacher assessment of students' oral performance in Filipino. Findings showed that teacher and students' self-assessment agree with each other. The study explains that high positive correlations confirm that both the students and the teacher consistently apply the rating scale. The author argues that the findings were due to the use of a micro-analytic rating scale, assessment training, and rating procedure used during the self-assessment.

Considering the above grounds, the present study aimed to investigate the differences between learners and teachers as assessors in the EFL writing context. The focus was on the probable mismatches between teachers' and students' criteria for assessing writing as well as investigating the effect that introducing standardized criteria would have on the reduction of mismatches between the students' and teachers' perception of what is important in self-assessing their writing performance.

This study primarily aimed to investigate the students' perception of writing assessment criteria by asking students to write on a particular topic and then assess their own works based on their perceived criteria during the writing instruction sessions. The next step was aimed at tailoring a standardized rubric by teachers' collaboration with the students. Therefore, by introducing the standardized criteria it attempted to identify if there was any significant improvement in the learners' performance in writing assessment. The study further explored whether learners' assessment of their writing agrees with that of teachers before and after being provided with the rating criteria. To achieve the above goals, the following research questions were posed in this study:

1. What criteria do learners use to evaluate their own writing performance before receiving the rating criteria?
2. Is there any significant difference in learners' assessment of their own writing before and after receiving the rating criteria?
3. Is there any relationship between learners' and teachers' assessment of writing before and after receiving the rating criteria?

### **3. Method**

#### ***3.1. Design of the Study***

The study adopted a mixed-method design since both qualitative and quantitative data collection and analysis procedures were used. Therefore, quantitative data including writing scores were collected from both teachers' and students' ratings, and qualitative data including students' comments on their own writing performance were used to probe into the criteria they used when rating their texts.

### 3.2. Participants

A body of six EFL teachers and 27 EFL students participated in this study. The teachers were all female and their age ranged between 26 and 35. Moreover, they had several years of experience in teaching and assessing the English language. All the teachers worked in the Iranian English Institute in Kangan, a town located in the south-west of Iran. Demographic information of the teachers is shown in Table 1. Henceforth, the students are referred to by pseudonyms (L1, L2, etc.) to maintain their anonymity in this study.

In addition, twenty-seven female students participated in this study. Their ages ranged between 13 and 23. The students were from 6 intact classes of intermediate learners. They had been assigned to this particular proficiency level following the institute assessment procedures. The course covered all four skills of the language including writing. Although the teachers followed the institute assessment procedures, they were free to devise their own formative assessments. A brief analysis by the researchers showed that the learners did not have any familiarity with the self-assessment procedure.

Table 1

*Demographic Information of the Teacher Participants in the Study*

	Gender	Educational background	Levels taught	Teaching context	Teaching experience (years)
T1	Female	MA in TEFL	Beginner/Intermediate	Language institutes	6
T2	Female	BA in TEFL	Beginner/Intermediate	Language institutes	9
T3	Female	BA in TEFL	Intermediate/advanced	Language institutes	10
T4	Female	MA in TEFL	Beginner/Intermediate	Language institutes	8
T5	Female	MA in TEFL	Intermediate/advanced	Language institutes/school	12
T6	Female	BA in TEFL	Intermediate/advanced	Language institutes	15

### 3.3. Instruments

The following instruments were used in this study.

**3.3.1. Writing Texts.** After a brief pilot phase, which showed that the students preferred argumentative essay topics, they were asked to write on the paper sheets prepared by the researchers and to provide information such as gender, major, and the name of their instructor. In addition, the students were provided with the required information for writing their text including the prompt and the needed details (the length of the text and the time limit). The texts collected in this way were used in the next self-assessment phase by the researchers.

**3.3.2. Writing Assessment Sheet.** The researchers developed a writing assessment sheet for the students to assess their texts. In the first phase of assessment, the students were required to evaluate their texts based on their own rating criteria. In addition to mentioning their rating criteria, the students were required to give a total score to their texts as well. In the second phase of the assessment, the students were provided with the same assessment sheets to evaluate their texts using their criteria. They were also required to give a total score out of 40 to the texts.

**3.3.3. Researcher-developed Rating Scale.** After the teacher participants in this study agreed on their final list of rating criteria, the researchers developed a scale which comprised eight components on a five-point Likert scale. The rating scale yielded a total score of 40. The researchers used the agreed-upon rating criteria in the training session. The

scale was used in the second self-assessment phase by the students. Table 2 shows the rating scale developed based on the teachers' stated rating criteria.

**3.4. Data Collection Procedure**

Due to the centrality of writing performance in this study, the researchers conducted a brief pilot study to identify the appropriate topic, length of the essay, and the time to write the essay. In piloting the writing text, researchers asked the students to choose and write about one of the three topics in three genres of description, argumentation, or narration. The analysis of texts showed that the majority of the students (70%) had selected the argumentative topic. Moreover, the analysis of the texts showed that 80% of the texts had a three-paragraph length. Although there were no time limitations, student writers finished their texts within 30-45 minutes. Therefore, a forty-five-minute time frame was considered an appropriate time for writing a three-paragraph essay. As a result of the pilot phase, the researchers decided on a three-paragraph argumentative essay to be written in 45 minutes. Next, the researchers met the students in their classes and asked them to write an argumentative essay on this topic, *social media is harmful for students* in 45 minutes. The writing task was prepared on a sheet. It included the prompt and the relevant instructions for specifying what the students were supposed to consider in their text.

Upon the completion of the writing task, the students were given a writing assessment sheet which asked them to write their own criteria for evaluating their writing. Moreover, the students were asked to read their texts and assign a score out of 30 to their texts.

The researchers, then, asked the teachers to develop a list of criteria based on which to assess the writing performance of the students. The teachers who had developed expertise in teaching and assessing different language skills, including writing, could consult available rating scales to develop their list of criteria (Jacobs et al.,1981). After developing their list of criteria individually, the teachers were asked to meet and consult other teachers to finalize their list. The researchers used the teachers' list of criteria to develop a rating scale which included 6 components on a 5-point Likert scale, yielding a total of 30. The following table shows the final scale developed by the researchers.

Table 2  
*Rating Scale Developed by Researchers in the Study*

Components	Content	Organization	Language use	Mechanics	Vocabulary	Cohesion & Coherence
Description	Quality of ideas	Essay development (paragraph development/structure)	Correct use of tense, number, prepositions & articles	Spelling, punctuation, capitalization, handwriting	Word choice	Relevance of ideas/sentences to each other
	Novelty of ideas	Rhetorical organization (thesis statement, organization development, structure of paragraph)	Sentence structure		Diction (choice of right lexical items)	Logical relationships among sentences

Next, to ensure the reliability of the developed scale and to discuss it, they were asked to rate a single text based on the scale. The results of the rating were discussed to decrease the discrepancies among the different raters. Then, the raters were given another text for another rating to ensure that they equally interpreted the scale and the way to use it.

Although the students did not have any experience with writing self-assessments, it was possible for other learning effects to contaminate the results. Therefore, to have an optimal rating between the two rating occasions, the researchers met the students after one month. The researcher shared the rating scale with the students. Moreover, they explained the rating criteria in the scale to them. In addition, to help students how to apply the rating scale, five samples of the students' essays were scored using the scale. The students were asked to raise their questions during the sample ratings by the researchers. As the researchers proceeded with 4<sup>th</sup> and 5<sup>th</sup> sample scoring, the students adopted a more active role and helped the researchers with their scoring. After conducting five sample ratings and clarifying the students' ambiguities regarding the scale, the students were provided with their texts and were asked to assess their writing using the scale. In addition to evaluating their texts based on individual criteria, the students were supposed to assign their text a score out of 40. Moreover, the teacher-participants in this study were provided with the students' texts and were asked to rate the texts based on the rating scale. They assessed the students' texts while not knowing about the students' self-assessments.

### 3.5. Data analysis

To answer the first research question, the criteria the students used for rating their texts on the first occasion were analyzed inductively and grouped into categories. Moreover, the two sets of quantitative data collected on two occasions regarding the second question were, also, entered into the SPSS program and were analyzed by the paired-samples t-test to discover whether any significant changes occurred as students were provided by the new agreed-upon scoring checklist.

As to the third question, in order to investigate any significant relationship between the sets of scores assigned to writing ability by learners themselves and the teachers on the two occasions, the Pearson product-moment correlation was run.

## 4. Results

### 4.1. The first research question

The first research question aimed to examine the details of the criteria mentioned by the learners before receiving the teachers' criteria and the training. The learners' comments on their own writing performance on the first occasion were inductively analyzed to find out what criteria they used to evaluate their own works. The result of the analysis showed that a considerable number of the learners mentioned language use, mechanics, and organization in assessing their own works. A summary of the frequency and percentage of each component is reported in Table 3 below.

Table 3

#### *Frequency and Percentage of Language Use, Mechanics, and Organization Mentioned by the Learners*

Categories	Frequency	Percentage (%)	Total
Mechanics	27	100.0	27
Language use	27	100.0	27
Organization	26	96.2	26

**4.1.1. Mechanics.** When evaluating their texts, all the students referred to some aspects of mechanics. These included spelling, capitalization, punctuation, and handwriting. The following shows how the students commented on different sub-components of mechanics in their writing.

**4.1.1.1. Spelling.** Spelling which refers to the correct dictation of the words in the text was similarly present in all the students' comments ( $f = 27$ ). For example, L1 wrote *I spell[ed] the words correctly.*



**4.1.1.2. Capitalization.** Similar to spelling, this criterion which refers to the adequate usage of capital letters at the beginning of the words, recurred in almost all the students' comments. For instance, L2, L4, L16, and L26 commented as so. *L2: I start sentences and proper nouns with capital letters; L4: I write capital letters correctly; L16: Sometimes I don't write capital [letters]; L26: I start[ed] every sentence with capital [letter].*

**4.1.1.3. Punctuation.** This aspect of mechanics which is concerned with using language marks in an appropriate way was considered important by a considerable number of the participants (f = 26). As an example of using this criterion, L8 stated *I used punctuation marks like period and question marks correctly. I always check them when revising my drafts.*

**4.1.1.4. Handwriting.** The analysis of the students' comments revealed that for a considerable number of the students, handwriting was also a part of mechanics. For example, L6 said *My handwriting is not bad and I think people can read it.*

**4.1.2. Language Use.** Although almost all of the students mentioned this category (f = 27) equally, they commented quite generally on the concept of grammar. They have not been able to go further into details and take further detailed information on aspects of language use such as word order, articles, pronouns, and prepositions into account. For instance, L1, L2, and L5 referred to their conception of grammar as follows: *L1: I think I didn't write the grammar perfectly; L2: I have correct grammar; L5: I don't have a very good grammar; I sometimes use wrong words.*

**4.1.3. Organization.** Regarding the organization of their essay, although the students could not comment on different aspects of the organization, they could more or less refer to some general aspects of it. For example, some of them mentioned this category in a more detailed way; while others pointed to a general idea about it. Based on the analysis of the learners' comments, this category was divided into three subcategories including the main idea, supporting ideas, and conclusion.

**4.1.3.1. Main Idea.** A considerable number of the students considered this subcategory as important in their writing self-assessment. As an example, L11 stated *I think my paragraph doesn't have a topic.*

**4.1.3.2. Supporting sentences.** A body of 17 learners referred to the subcategory of supporting ideas in their scoring criteria. For instance, L8 and L12 mentioned *L8: I have ....and supporting sentence,* and *L12: I have supporting sentence [es],* respectively.

**4.1.3.3. Conclusions.** Learners also showed awareness of the conclusion part of their assessment. For example, L26 stated as, *My paragraph has ....and conclusion*

## 4.2. The Second Research Question

To address the second research question, a paired-sample t-test was employed to investigate the effect of writing scoring criteria on the learners' ratings. A comparison between the two sets of mean scores showed that the learners gave themselves a slightly higher score in the first rating occasion than the second one (Table 4).

Table 4  
*Results of Paired-Samples Descriptive Statistics for the Students' Scores in the Two Rating Occasions*

	Mean	N	SD	SEM
Before the criteria	34.84	27	4.72	.92
After the criteria	32.50	27	5.1	1.001

Table 4 presents the results of the paired samples t-test. The paired-samples t-test indicated that scores were not significantly higher for the learners' rating in the first time ( $M = 34.8$ ,  $SD = 4.72$ ) than for their ratings in the second time ( $M = 32.5$ ,  $SD = 5.10$ );  $t(25) = 1.86$ ,  $p < 0.05$ .

Table 5  
*Results of the Paired-samples T-test Between the Students' Scores Before and After Receiving the Rating Scale*

M	SD	SEM	95% CI	t	df	Sig.(2-tailed)
1.34	3.67	.72	[.13, 2.83]	1.86	26	.024

**4.3. The Third Research Question**

The third research question examined the extent of agreement between the students' and the teachers' assessments before and after the learners were provided with the criteria and the training session. Therefore, to calculate the correlations between teacher assessment and the students' self-assessment on the two occasions, Pearson product-moment correlation coefficient was used. As Table 6 shows, the correlation between the learners' and the teachers' assessment on the first occasion was not statistically significant, as  $r(27) = 0.398, P > 0.05$ . On the other hand, in the second phase of the study, there was a positive correlation between the two sets of scores ( $r = .647, p < 0.05$ ). According to the concept of effect size introduced by Cohen (1988), a Pearson correlation coefficient larger than 0.50 is considered to be a large one; therefore, it can be concluded that there was a strong agreement between the teachers' assessment and the students' assessment after the students' receiving the rating scale ( $r = 0.64$ ).

Table 6  
*Pearson Product-moment Correlation Between Teachers' and Learners' Assessment*

	Teachers' assessment	Students' scoring before training	Students' scoring after training
Teachers' assessment	Pearson Correlation	1	.398
	Sig. (2-tailed)		.140
	N	27	27

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

**5. Discussion**

The findings of the study indicated that a number of factors were involved in the students' and teachers' mentioned criteria when rating writing. An analysis of the learners' comments on their own writing before being provided with the agreed-upon criteria showed that they were mainly concerned with language use, mechanics, and organization.

A more detailed analysis of these criteria showed that the students had a surface interpretation of these criteria. In fact, they considered grammar, spelling, capitalization, punctuation, main ideas, and supporting ideas. In other words, the students failed to consider all the characteristic components of the macro-level aspects of the text (Yu, 2007). In addition, they did not refer to significant components involved in writing assessment such as content, language structure, and vocabulary.

A further look into the students' self-mentioned criteria revealed that they were not able to judge their texts prior to being provided with the assessment criteria. The poor students' conceptualization of the writing criteria in the Iranian EFL context is due to several reasons. First of all, writing assessment has remained a mysterious task in the context. The students usually do not receive detailed feedback on their performance. In a product-oriented writing instruction context, the students are usually provided with a final score; therefore, they cannot locate the important aspects of the text. According to Sadler (1989), providing learners with generalities and abstract notions in teaching adversely affects their learning; consequently, they cannot translate what they have been taught in their assessment of their work.

To add more complexity, many studies (e.g. Ghanbari & Barati, 2014) have shown that Iranian EFL teachers do not use any explicit rating scales in their rating and they mostly rely on their impressions

when assessing the learners' texts. Along these, the students' lack of familiarity with the self-assessment procedure also negatively affected the students' self-awarded ratings (Orsmond, et al., 2000; Polanyi, 1958). Putting together these factors, it can be found why the students in this study counted some limited aspects of writing.

Moreover, the findings showed a significant difference between the students' assessment before and after being provided with the assessment criteria and the practice session. In the first round of scoring, the students experienced noticeable dissimilarities. As mentioned above, they had not received any systematic feedback on their writing, so they did not have a clear conceptualization of the writing construct. In the same line, they were not familiar with the self-assessment procedure. After receiving the teachers' agreed-upon assessment criteria, the students got familiar with different components of writing. Moreover, they learned how to apply those criteria in their assessment.

The analysis of the students' self-rating showed that they assigned higher scores to their writing on the first occasion. It can be argued that the students overestimated their writing ability simply because they were assessing their texts according to their own criteria. They assessed their writing while they were not aware of the multi-faceted nature of the writing construct. Their self-mentioned criteria revealed that they founded their assessment based on general and superficial aspects of the writing (Ghanbari, 2019).

The students' self-ratings in the second phase demonstrated that when students as important stakeholders participate in the assessment process, they develop a more consistent concept of the assessment task and the construct under assessment. In a relevant study, Ghanbari (2019) showed when test-takers become informed of the rating criteria and the assessment procedure, their performance considerably improves. When the students are involved in the rating, they learn about the rating criteria and negotiate them with the teachers. In this collaborative environment, assessment becomes more humanistic. Here, the test-takers are no more passive instruments who should receive some particular forms of knowledge and then be assessed on some vague criteria. When the assessment criteria and the processes and procedures of assessment are made explicit, the practice of self-assessment reduces the perceptual gap between the students and the teachers. As a result, there will be more consistency between what teachers intend and what the students interpret (Kumaravadivelu, 1991).

According to Norton and Starfield (1997), accountability in language assessment improves when the test-takers know about the assessment criteria. Yu (2007) involved the test-takers in developing a scoring checklist to evaluate their written summaries in order to reduce the perceptual mismatches between the testers and the test-takers. In the same line, some studies have shown that learners' viewing of a scoring rubric prior to completing a writing task can help them to write better (Howell, 2011; Sundeen, 2014). Similarly, some studies have emphasized that including the students in developing rating rubrics helps them to become actively engaged in developing the criteria for their own assessment (Panadero & Johnson, 2013; Skillings & Ferrell, 2000; Sundeen, 2014). The common thread of all these studies is emphasizing the active involvement of the students in their assessment in order to make the assessment more ethical, democratic, and fair (Lynch, 2001; McNamara, 2001; Shohamy, 2001).

From a more macro-level perspective, the enhanced performance of the students in the second phase of the study proved democratic approaches to assessment. By involving the students in their assessment, the power relations between students as test-takers and teachers as raters are redressed. In fact, sharing the power with the students turns the assessment task into a meaningful activity which significantly enhances assessment validity and fairness (Bachman & Palmer, 1996; Yu, 2007).

Regarding the third research question, the results showed that providing the students with the rating criteria and practice sessions would cause greater consistency between their self-assessment and teacher rating. This supports the findings of the previous studies that mentioned the benefits of involving the students in metacognitive activities. By including the students in the self-assessment activities, learners' understanding of what teachers aim to say improves, and consequently the mismatches between the students and the teachers reduce (Andrade et al., 2009; Babaii et al., 2016; Becker, 2016; Li & Lindsey, 2015; Patri, 2002; Perkins, 1992; Skillings & Ferrell, 2000; Saito & Fujita, 2004; Trace et al., 2017; Xiao, 2007).

According to Kumaravadivelu (2003), perceptual mismatches between teaching objectives and learning objectives always exist. A good number of studies (Babaii, et al., 2016; Barkhuizen, 1998; Block, 1994, 1996; Dehghan Harati, 2011; Hawkey, 2006; Lee & Choi, 2019; Kamali & Behjat, 2018; Kumaravadivelu, 2003; Ramazani, 2014; Slimani, 1989; Sengsouliya et al. (2021); Tajgozari & Alimorad, 2019; Tragant, 1997) has shown that teachers and students have different interpretation of the instruction. The innovative practice of self-assessment along with the students' poor familiarity with the constituting components of the writing construct deepened the perceptual gap between the students and the teachers. However, when the students were provided with the writing assessment criteria and learned how to seek them in their writing tasks, more convergence appeared between their self-rating and teachers' assessment.

In addition to reducing the perceptual mismatch between the teachers and the students, the findings revealed that involving students in classroom assessment tasks would help teachers get closer to the students' cognitions and conceptualizations of what has been instructed. The greater proximity between teachers and students would not only reduce the number of mismatches in the classroom but also by identifying the gaps, learning opportunities can be designed to address them purposively, and as a result, there would be a greater coherence between what teachers intend to teach and how the students interpret them which in turn increases the integration in the perception of learning and teaching objectives between the students and their teachers in the long run.

Although assessment reforms have been considered by many countries, assessment practice in many of these countries still lacks a significant change (Buhagiar, 2007). The possible reasons are believed to be caused by weaknesses in classroom evaluation practice which put an emphasis on superficial and rote learning and recall of isolated ideas which the student soon forgets. Other reasons are the overemphasis on grading function and paying less attention to the learning process on the one hand, and the preference for competition among the students over personal improvement on the other hand (Black & Williams, 1998).

Assessment will truly support learning; therefore, if the students are involved in their learning process (e.g. self-assessment) it will complement the teacher assessment. However, the issue is that unless the capacity of self-assessment is developed and standards of quality are absorbed, the students will not be independent and autonomous learners (Black & Williams, 1998; Sadler, 1989; Stefani, 1994). On the other hand, the teacher should step out of a provider of assessment information role and move towards guiding and consulting the students by working in collaboration so that the students gain a better insight into the learning goals (Buhagiar, 2007; Stefani, 1994). Cross (1998) agrees that in such contexts, students will be able to make a critical self-judge as well as assessing their peers. Moreover, they can be autonomous to carry on their goals even if they are out of formal education contexts.

## 6. Conclusion

The findings of the present study offer several implications. First of all, providing learners with an opportunity for self-assessment causes the students to have a more dynamic role in assessing their own works. Besides, it gives them the opportunity to focus more on the important aspects of assessing writing scripts and recall what they have learned, and put it into practice as well. On the other hand, this enables the teachers to have a better insight into how their students perceived the training they have been provided with by means of negotiation and feedback on what they have covered in their classrooms. The results recommend moving towards a more integrated curriculum by identifying the gaps and training the students in the areas of problem.

Providing agreed-upon criteria in training sessions can also lead to reducing the mismatches and establishing a common understanding of the assessment criteria between the teacher and the learners (Manzano, 2022; Orsmond et al., 2000). By investigating the differences between the learners' and teachers scoring in the first and second time, the effect of negotiating and training on how to assess writing performance, and consequently students' better understanding of what teachers mean and the reduction of mismatches will be noticeable.

Moreover, as a practical implication of this study, instructors can provide learners with the opportunity to constantly monitor and improve their learning through participating in the ongoing procedure of self-assessment. In this way, the learners can have ownership of their successes and failures. As a further

contribution, developing and sharing the rating criteria with the learners develop and maintain a sense of accountability for all major ELT stakeholders (e.g. policy-makers, syllabus designers, curriculum developers, and teacher trainers).

The present study also faced some limitations. As the first constraint, the use of 27 Iranian EFL students who were from six intact classes and a group of six teachers in a particular context hindered the generalizability of the findings of the study. Future studies should be conducted with a larger and more representative sample of students and teachers in different contexts. Moreover, students' proficiency level was not taken into account in this study and the researchers only relied on the institute placement of the students into the intermediate level. Regarding the importance of the student's proficiency level in the reliability of their self-assessment (Shimura, 2006), the proficiency level of the students should be considered in future studies on self-assessment. Moreover, all the participants in the study (i.e. teachers and students) were female which might have introduced some degrees of bias into the study. Including both gender types in future studies is recommended. Last, although the students mentioned their rating criteria in the first scoring occasion, adopting more qualitative measures such as think-aloud protocols can shed light on their conceptualization of the writing construct and the self-assessment practice.

This study also recommends several areas for future studies. The focus of the present study was on self-assessment as one technique of alternative assessment in the classroom. Further studies can direct the study toward other alternative assessment procedures such as the peer assessment to gain a better understanding of other aspects of assessment. The effect of gender in this study was not taken into account. In order to have a better insight into the effect of this important factor, future research can pay attention to the differences in rating among learner participants of different genders. The study investigated the research questions among intermediate learners. It is suggested that investigating the learners with higher levels of language proficiency might have different results and enrich the literature.

Further research can focus on other genres of writing. In the current study, the learners were asked to write an argumentative text. Future research can use other genres (e.g. narrative, explanatory, descriptive) of writing. They can also investigate more than one genre to gain more helpful results from other aspects of the research.

As the result revealed, providing students with criteria might not be sufficient for them to be autonomous raters with a desirable amount of knowledge in different areas of assessment. This study calls for further investigation on involving students in constructing assessment criteria and its influence on their performance when rating their own work.

**Declaration of Conflicting Interests:** The authors confirm that there is no conflict of interest to declare.

**Funding:** There are no sources of funding for the research study.

## References

- Andrade, H., Buff, C., Terry, J., Erano, M., & Paolino, S. (2009). Assessment-driven improvements in middle school students' writing. *Middle School Journal*, 40(4), 4-12. <https://files.eric.ed.gov/fulltext/EJ833638.pdf>
- Arnold, J. (2022). Prioritizing students in assessment for learning: A scoping review of research on students' classroom experience. *Review of Education*, 10, 1–16. <https://doi.org/10.1002/rev3.3366>
- Babaii, E., Taghaddomi, sh., & Pashmforoosh, R. (2016). Speaking self-assessment: Mismatches between learners' and teachers' criteria. *Language Testing*, 33(3), 411-437. <https://doi.org/10.1177/0265532215590847>
- Bachman, L., & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Barkhuizen, G. P. (1998). Discovering learners' perceptions of ESL classroom teaching/learning activities in a South African context. *TESOL Quarterly*, 32(1), 85-108. <https://doi.org/10.2307/3587903>
- Becker, A. (2016). Student-generated scoring rubrics: Examining their formative value for improving ESL students' writing performance. *Assessing Writing*, 29, 15-24. <https://doi.org/10.1016/j.asw.2016.05.002>

- Black, P., & Williams, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-148. <https://kappanonline.org/inside-the-black-box-raising-standards-through-classroom-assessment/>
- Block, D. (1994). A day in the life of class: Teachers/learner perceptions of task purpose in conflict. *System*, 22(4), 473-486. [https://doi.org/10.1016/0346-251X\(94\)90004-3](https://doi.org/10.1016/0346-251X(94)90004-3)
- Block, D. (1996). A window on the classroom: Classroom events viewed from different angles. In K. M. Bailey & D. Nunan (Eds.), *Voices from the classroom: Qualitative Research in Language Classrooms* (pp. 168-193). Cambridge: Cambridge University Press. [https://doi.org/10.1016/0346-251X\(94\)90004-3](https://doi.org/10.1016/0346-251X(94)90004-3)
- Boud, D. (1995). *Enhancing learning through self-assessment*. London: Kogan Page.
- Boud, D. (2013). *Enhancing learning through self-assessment*. New York: Routledge.
- Bourke, R., & Mentis, M. (2013). Self-assessment as a process for inclusion. *International Journal of Inclusive Education*, 17(8), 854-867. <https://doi.org/10.1080/13603116.2011.602288>
- Brown, A. V. (2006). *Students' and teachers' perceptions of effective teaching in the foreign language classroom: A comparison of ideals and ratings*. [Unpublished M.A. thesis]. University of Arizona, USA.
- Buhagiar, M. (2007). Classroom assessment within the alternative assessment paradigm: Revisiting the territory. *The Curriculum Journal*, 18(1), 39-56. <https://doi.org/10.1080/09585170701292174>
- Cohen, J. W. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Dehghan Harati, R. (2011). Learners' beliefs, teachers' beliefs: Are they different? *The Iranian EFL Journal*, 7(5), 190-221.
- Eun, H. (2015). A study on perceptions of feedback between students and teachers in English writing course. *Studies in English Language & Literature*, 41(3), 203-232. <https://doi.org/10.5782/2223-2621.2019.22.4.128>
- Falchikov, N. (2013). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education*. New York: Routledge.
- Ghanbari, N. (2019). Promoting fairness in EFL writing assessment: Are there any effects of the writers' awareness of the rating criteria? *The Journal of Asia TEFL*, 16(2), 448-767. <http://dx.doi.org/10.18823/asiatefl.2019.16.2.22.735>
- Ghanbari, N. & Barati, H. (2014). Iranian EFL writing assessment: The agency of rater or rating scale? *International Journal of Language Testing*, 4(2), 204-228. [https://www.ijlt.ir/article\\_114401.html](https://www.ijlt.ir/article_114401.html)
- Harper, J. & Puji Widodo, H. (2020). Perceptual mismatches in the interpretation of task-based ELT materials: A micro-evaluation of a task-based English lesson. *Innovation in Language Learning and Teaching*, 14(2), 114-132. <http://dx.doi.org/10.1080/17501229.2018.1502773>
- Howell, R. J. (2011). Exploring the impact of grading rubric on academic performance: Findings from a quasi-experimental pre-post evaluation. *Journal of Excellence in College Teaching*, 22, 31-49.
- Jacobs, H. L., Zingraf, S. A., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). *Testing ESL composition: A practical approach*. Rowley, MA: Newbury House
- Johnson, R. K. (1989). *The second language curriculum*. Cambridge: Cambridge University Press.
- Kamali, M., & Behjat, F. (2018). Types of perceptual mismatches in intermediate EFL classes. *Journal of Applied Linguistics and Language Research*, 5(3), 156-168. <http://www.jallr.com/index.php/JALLR/article/view/834>
- Kumaravadivelu, B. (1991). Language-learning tasks: Teacher intention and learner interpretation. *ELT Journal*, 45(2), 98-107. <https://doi.org/10.1093/elt/45.2.98>
- Kumaravadivelu, B. (2003). *Beyond Methods: Macrostrategies for language learning*: New Haven: Yale University Press.
- Lee, J. & Choi, E. (2019). Mismatches between student and teacher perceptions of current and ideal foreign language classroom practices. *Korean Journal of Applied Linguistics*, 35(3), 77-102. <http://dx.doi.org/10.17154/kjal.2019.9.35.3.77>
- Li, J., & Lindsey, P. (2015). Understanding variations between student and teacher application of rubrics.

- Assessing Writing*, 26, 67-79. <https://doi.org/10.1016/j.asw.2015.07.003>
- Luoma, S. Y., & Tarnanen, M. (2003). Creating a self-rating instrument for second language writing: From idea to implementation. *Language Testing*, 20(4), 440-465. <https://doi.org/10.1191/0265532203lt267oa>
- Lynch, B. K. (2001). Rethinking assessment from a critical perspective. *Language Testing*, 18(4), 351-372. <https://doi.org/10.1177/02655322010180040>
- Manzano, D. L. (2022). Examining the interrater reliability between self- and teacher assessment of students' oral performances. *International Journal of Language Testing*, 12 (2), 128-144. [10.22034/ijlt.2022.157130](https://doi.org/10.22034/ijlt.2022.157130)
- Manzari, A. (2023). Reliability and validity of self-assessments among Iranian EFL university students. *International Journal of Language Testing*, 13(1), 18-43. [10.22034/ijlt.2023.363420.1199](https://doi.org/10.22034/ijlt.2023.363420.1199)
- Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing*, 26(1), 075–100. <https://doi.org/10.1177/0265532208097337>
- McNamara, T. (2001). Language assessment as social practice: Challenges for research. *Language Testing*, 18(4), 333-349. <https://doi.org/10.1177/026553220101800402>
- Montgomery, J. L., & Baker, W. (2007). Teacher-written feedback: Student perceptions, teacher self-assessment, and actual teacher performance. *Journal of Second Language Writing*, 16, 82-99. <https://doi.org/10.1016/j.jslw.2007.04.002>
- 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. <https://doi.org/10.1080/03075070600572090>
- Norton, B., & Starfield, S. (1997). Covert language assessment in academic writing. *Language Testing*, 14(3), 278-294. <https://doi.org/10.1177/026553229701400305>
- Orsmond, P. (2011). *Self- and peer-assessment: Guidance on practice in the biosciences*. Leeds: The Higher Education Academy Center for Biosciences.
- Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education*, 25(1), 23-38. <https://doi.org/10.1080/02602930050025006>
- Panadero, E., & Johnson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited. *Educational Research Review*, 9, 129-144. <https://doi.org/10.1016/j.edurev.2013.01.002>
- Patri, M. (2002). The influence of peer feedback on self- and peer-assessment of oral skills. *Language Testing*, 19(2), 109–131. <https://doi.org/10.1191/0265532202lt224oa>
- Perkins, D. (1992). *Smart school: From training memories to educating minds*. New York, NY: The Free Press.
- Polanyi, M. (1958). *The Study of Man*. London: Routledge & Kegan Paul.
- Ramazani, M. (2014). Mismatches in beliefs between teachers and students, and characteristics of effective English teacher: An Iranian context. *Procedia - Social and Behavioral Sciences*, 98, 1518 – 1527. <https://doi.org/10.1016/j.sbspro.2014.03.573>
- Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144. <https://doi.org/10.1007/BF00117714>
- Saito, H., & Fujita, T. (2004). Characteristics and user acceptance of peer rating in EFL writing classrooms. *Language Teaching Research*, 8(1), 31–54. <https://doi.org/10.1191/1362168804lr133oa>
- Sengsouliya, S., Soukhavong, S., Phonekeo, S., Sengsouliya, S., & Xaixanith, T. (2021). Mismatches in teachers' teaching and students' learning styles in English classes at a secondary school level: A case study of Laotian secondary schools. *International Journal of Research in English Education*, 6(1), 16-33. <https://doi.org/10.52547/ijree.6.1.16>
- Shimura, M. (2006). Peer- and instructor assessment of oral presentations in Japanese university EFL classrooms: A pilot study. *Waseda Global Forum*, 3, 99–107.

- <https://core.ac.uk/download/pdf/286942097.pdf>
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language test*. Harlow, England: Longman.
- Skillings, M. J., & Ferrell, R. (2000). Student-generated rubrics: Bringing students into the assessment process. *The Reading Teacher*, 53, 452–455. <http://www.jstor.org/stable/20204821>
- Slimani, A. (1989). The role of topicalization in classroom language learning. *System*, 17(2), 223-234. [https://doi.org/10.1016/0346-251X\(89\)90035-3](https://doi.org/10.1016/0346-251X(89)90035-3)
- Somervell, H. (1993). Issues in assessment, enterprise and higher education: The case for self/peer and collaborative assessment. *Assessment & Evaluation in Higher Education*, 18(3), 221–233. <https://doi.org/10.1080/0260293930180306>
- Sundeen, T. H. (2014). Instructional rubrics: Effects of presentation options on writing quality. *Assessing Writing*, 21, 74-88. <https://doi.org/10.1016/j.asw.2014.03.003>
- Stefani, L. A. J. (1994). Peer, self and tutor assessment: Relative reliabilities. *Studies in Higher Education*, 19(1), 69-75. <https://doi.org/10.1080/03075079412331382153>
- Tajgozari, M., & Alimorad, Z. (2019). Examining the possible effects of (mis)matches between EFL teachers' and students' perceptions of L2 writing assessment on students' writing achievement scores. *Khazar Journal of Humanities and Social Sciences*, 22(4), 128-145. <https://doi.org/10.5782/2223-2621.2019.22.4.128>
- Trace, J., Janssen, G., & Meier, V. (2017). Measuring the impact of rater negotiation in writing performance assessment. *Language Testing*, 34(1), 3-22. <https://doi.org/10.1177/0265532215594830>
- Tragant, E. (1997). Mismatches between teachers' expectations and students' performances. *Revista Alicantina de Estudios Ingles*, 10, 193-208. [https://rua.ua.es/dspace/bitstream/10045/6003/1/RAEI\\_10\\_14.pdf](https://rua.ua.es/dspace/bitstream/10045/6003/1/RAEI_10_14.pdf)
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Williams, E. (1992). Student attitudes towards approaches to learning and assessment. *Assessment and Evaluation in Higher Education*, 17(1), 45-58. <https://doi.org/10.1080/0260293920170105>
- Wolsey, T. D., Lapp, D., & Fisher, D. (2012). Students' and teachers' perceptions: An inquiry into academic writing. *Journal of Adolescent & Adult Literacy*, 55(8), 714–724. <https://doi.org/10.1002/JAAL.00086>
- Xiao, Y. (2007). Applying metacognition in EFL writing instruction in China. *Reflections on English Language Teaching*, 6(1), 19-33. <https://www.nus.edu.sg/cehc/wp-content/uploads/2022/11/19-34xiao.pdf>
- Yu, G. (2007). Students' voices in the evaluation of their own written summaries: Empowerment and democracy for test takers. *Language Testing*, 24(4), 539-572. <https://doi.org/10.1177/0265532207080780>