

An Inter-Rater Perspective for the Researches on Assessing Writing

Turgay HAN

Kafkas University, Faculty of Science and Letters,
Department of English Language and Literature, Kars, Turkey
turgayhan@yahoo.com.tr

Hüseyin EFE

Atatürk University, Faculty of Letters,
Department of English Language and Literature, Erzurum, Turkey
hefe@atauni.edu.tr

Erdinç Parlak

Atatürk University, Kazım Karabekir Faculty of Education,
Department of English Language Education, Erzurum, Turkey
erdincparlak@hotmail.com

Abstract: In assessing EFL students' writings consistently, there are various factors that must be taken into consideration while rating. Especially those who want to make quantitative researches using raters or who aim to give suitable feedback to written productions should be responsible for fulfilling the requirements in marking and scoring process. In this context, this paper looks insight for some issues related to the raters, such as inter-rater reliability, analytic or holistic examinations, rating criteria, and others. This study includes 8 native raters and 8 non-native raters, each of whom rated an ESL essay both holistically and analytically. Every participant rater's background of scoring ESL writings was similar. The result showed that there was no significant difference between raters, that is, participants' grading of the essay is irrespective of their being native speakers. At the end of the study, some important implications for essay rating practices and both the researchers and language teacher were emphasized.

Key Words: Rater, Reliability, Assessing Writing, EFL

Introduction

Many researches based on assessing writing and error treatment incline some variables with regard to the scoring each ESL writing accurately. In the process of deciding the effects of any types of feedback given to the EFL students' writings, there are some sorts of musts that researchers fulfil in the course of doing their statistical analyses related to the data obtained by their scorings methods. Even though several factors influence scoring and the process of raters' decision makings, researchers in the field of ESL essay rating delve into varying issues such as task requirement, rater characteristics and essay characteristics (Barkaoui, 2010; p.54). Assessing L2 writing accurately is very important for the validity of the inferences. Therefore, essays judged by more than one examiner will be closer to the fair score than judgement made by only one rater (Hamp-Lyons, 1990; p.79). Writing assessment process involves a multi-dimensional evaluation; for this reason, clarity, coherence and grammatical quality are some of the core points to be assessed for a writing paper. In this context, inter-rater reliability, one of the components of writing assessment process, is considerably the critical issue in scoring EFL/ESL writings as there are a few interfaces that raters are prone to experience such as the idiosyncratic, rating methods, and criteria during assessment. Since it is a subjective phenomenon, the decisions made through the scores given have some potentialities that effect overall research. In this research, to assure objectivity in scoring, two types of raters were chosen. As Stemler (2004) emphasizes, "Raters are often used when student products or performances cannot be scored objectively as right or wrong but require a rating degree. The use of raters results in the subjectivity that comes hand in hand with an interpretation of the product or performances (cited in Bresciani, Oakleaf, et al, 2009; p.3)".

Potentialities of Writing Assessment

Hyland (2003) states that there are four basic elements in the design of writing assessment: Rubric (the instructions), prompt (the task), expected response (what the teacher intends students to do with the task), post-task evaluation (assessing the effectiveness of the writing task) (cited in Coombe, Folse, & Hubley, 2007; p.71).

Sudweeks et al (2005) explain the factors involved in the rating a particular writing paper as external and internal ones. Sudweeks et al advocate that as students vary in their writing abilities it is difficult to expect them to receive the same rating or they cannot be measured the same, and the external factors can be counted for receiving rating are: (a) the nature of the particular writing prompt or task posed, (b) the raters who judge students' essays (c) situation-specific factors (d) students' background and interests related to the topic that will be developed (e) interactions among these varying sources (p.240).

Raters are the fundamental variables in the assessing process. According to Weigle (2002) there are two foci points in assessment: a consideration of what attributes of the compositions raters focus on and the effects of background rater characteristics and of the process of reading compositions and ultimately on the scores that raters use (p.70). Researchers in recent years have looked for the effects attributed to the raters on ratings. Sweedler-Brown (1985) realized that rater trainers were more unforgiving in their scorings when compared to the ratings of novice raters (cited in Weigle, 2002; p.71).

There are also some methods developed for analyzing writings accurately. For example, Generalizability Theory and Many-Facet Rasch Model are used in some researches to (a) estimate the sources causing rating errors, (b) to achieve reliability estimates, and (c) to make recommendations for improving rating process (Sudweeks, Reeve, & Bradshaw, 2005; p.239).

In this context, scoring involves defining the rating scale and ensuring that raters use the scale appropriately and consistently as two central considerations (Weigle, S.C. 2002, p.108).

Stemler (2004) explains that inter-rater reliability requires consensus estimates, consistency estimates and measurement estimates approaches to assure the accuracy and consistency in scoring. Consensus estimates come into meaning that the degree to which markers give the same scores. Consistency estimate is a sign of the extent of the similarities of the high and low scores among the markers. Additionally, measurement estimates show the degree to which scores can be attributed to common scoring rather than error components (cited in Brown, Glasswell, & Hardland, 2004; p.106).

Moreover, consistency in scoring requires two sub-processes to eliminate the sources causing unreliable scoring. Weigle(2002) indicates two main types of inconsistencies: "[...] (1) Inconsistencies in the ratings of a single scorer across different scripts of similar qualities or the same script on different occasions, and (2) inconsistencies between different scores..." (pp.128-129).

The Aspects in Rating

Two main aspects of reliability involve raters' markings which are the indications of consistencies of scorings.

Inter-Rater Reliability and Intra-Rater Reliability

Subjective marking of writing requires a measurement of reliability within a rater's marking and among raters. The term 'inter-rater reliability' refers to the self consistency and the extent of similarity of ratings of different markers, that is, scoring of a single writing of a rater has not significant difference when compared to the one of other independent raters (Weigle, 2002), whereas intra-rater reliability contrasts with inter-rater reliability. Bachman (1995) advocates that "Ratings given by different raters can also vary as a function of inconsistencies in the criteria used to rate and in the way in which these criteria are applied..." (p.180). Bachman (1995), further, explains that when several essays are given to some independent raters for their ratings, the results probably will not be the same. The variability of the criteria of the raters can be counted as the outstanding source for these inconsistencies as some raters may look for the quality of content and some other may look for the organization (p.180).

On the other hand, "[...]intra-rater reliability refers to the tendency of a rater to give the same score to the same script on different occasions, while inter-rater reliability refers to the tendency of different raters to give the same scores to the same script..." (Weigle, 2002; p.135).

Rationales for Holistic Scoring and Analytic Scoring

Making inferences from the scores given to writings is variable as it depends on the method of scoring. In this context, "[...]Traditionally a student's writing performance was judged in comparison with the performance

of others, but this *norm-referenced* method has largely given way to *criterion-referenced* practices where the quality of each essay is judged in its own right against some external criteria, such as coherence, grammatical accuracy, contextual appropriacy, and so on..." (Hyland, 2008; p.226).

Moreover, scales that will be used for rating should cover objective characteristics as possible. Therefore, two questions can be raised to inquire a system for scoring a script: whether to give a single score to each script, or score each script depending on several different features. Three main approaches consisting of *primary trait scales*, *holistic scales*, and *analytic scales* are available in the composition assessment literature. Holistic scoring, or general impression marking, means that giving single scores to the writing by judging an overall impression. On the other hand, analytic scoring depends on several aspects of writing rather than a single score (Weigle, 2002; pp. 64-115).

There are, of course, some probable limitations for both types of scoring. In holistic scoring, using the scale for analytic scoring bears some problems as it includes several sub-categories for rating. Then, again, the practicality of analytic scoring can be judged in terms of time consuming and difficulty of applying the criteria to each script (East, 2009; p.91). Furthermore, holistic scoring can be used to attract the attention of the readers to the strength of the scripts and not to the lack of it and thus, writers can be reinforced and awarded (White, 1984, 1985). Nevertheless, "analytic scoring schemes thus provide more detailed information about a test taker's performance in different aspect of writing and are for this reason preferred over holistic schemes by many specialist" (Weigle, 2002; pp.114-115).

There are some researches on the reliability of different scale types. In this context, Weir (1990) explains that analytic scoring is more reliable than holistic scoring (cited in Weigle, 2002, p.73). Also, Bauer (1981) found that even though holistic scoring is more cost effective, analytic scoring is more reliable than holistic scoring (cited in Weigle, 2002; p.73).

Additionally, in the scoring process, validity is another decisive factor referring to the writing course goals and outcomes. In order to ensure that writings are scored objectively and to obtain high inter-rater reliability, there are some internal and external factors in the setting for marking errors in compositions of subjects under questions. Therefore, as defined in the *Standards for Education and Psychological Testing*, "Validity... is a unitary concept. Although evidence may be accumulated in many ways, validity always refers to the degree to which that evidence supports the inferences that are made from the scores. The inferences regarding specific uses of a test are validated, not the test itself" (American Psychological Association, 1985, p.9). It must be noted that "...any type of assessment should reflect the course goals..." (Coombe, Folse, Hubley, 2007, p.70).

Background

In this study, scoring reliably the learners' scripts by the researchers gains the importance. Participants of the present study are 8 experienced and non-native English language teachers and 8 native English language teachers holding at least MA degree in language teaching.

A writing paper was randomly selected from the administered writing exams given to the first year students studying English Language and Literature during the academic year 2010, and the paper was assessed by those 16 language teachers through scoring at two levels: holistic and analytic. Scoring processes of the paper were done according to a predetermined rubric. The scores of the participants got from holistic scale and analytic scale were statistically analyzed. Two research questions are addressed:

1. Is there any difference between the holistic ratings and analytic ratings of the participants in terms of inter-rater reliability?
2. To what extent do the mean of the scores of those 8 non-native English language teachers and 8 native English language teachers deviate?

Methodology

In this study, the raters used a "Direct Measures of Writing Assessment" approach while scoring the papers. In the direct approach, in contrast to indirect approach, students' success in communicating through the written mode of texts is taken into consideration. This type of assessment process includes the quality of content, appropriate vocabulary, grammar and syntax (Coombe et al, p.71).

Scoring Holistic Rubric

The intermediate level ESL/EFL script was rated according to a pre-determined rubric for holistic scoring which includes 5 types of categories and in which general writing ability can be judged (*see figure 1*). Using such

a rubric, of course, has some advantages besides having disadvantages. Hyland (2003; p.227) compares and contrasts the gain and loss of Holistic scoring:

Table 3.1. Advantages and Disadvantages of Holistic Scoring

Advantages	Disadvantages
Global impression no a single ability	Provides no diagnostic information
Emphasis on achievement not deficiencies	Difficult to interpret composite score
Weight can be assigned to certain criteria	Smooths out different abilities in subskills
Encourages rater discussion and agreement	Raters may overlook subskills
	Penalizes attempts to use challenging forms
	Longer essays may get higher scores
	One source reduces reliability
	May confuse writing ability with language proficiency

The participants of the study, both 8 native raters and the other 8 non-native raters, scored the intermediate level script according to the holistic scale (see figure 1). The requirements of the settings for the raters, rater training, and others were supplied in the process of reading.

Scoring Analytic Rubric

The paper under question was also rated by the participant raters through using a multiple-trait scoring rubric (*see figure 2*). This type of rubric is considered to be an ideal compromise on account of requiring raters to provide separate scores for different writing features. Furthermore, it is like a multifaceted scoring and in particular contexts and purposes, such rubrics can address the traits that do not occur in more analytic score (Hyland, 2003, p.230).

Hyland (2003, p.230) explains the advantages and disadvantages of analytic scoring in the below table:

Table 3.2. Advantages and Disadvantages of Analytic Scoring

Advantages	Disadvantages
Encourages raters to address the same features	May divert attention from overall essay effect
Allows more diagnostic reporting	Rating one scale may influence others
Assists reliability as candidates gets several scores	Very time consuming compared with holistic method
Detailed criteria allow easier rater training	Writing is more than simply the sum of its parts
Prevents conflation of categories into one	Favors essays where scalable into easily extracted
Allows teachers to prioritize specific aspects	Descriptors may overlap or ambiguous

Obtaining High Inter-rater Reliability

In this study, to keep raters reliable in their scoring marks there are some requirements that should be applied in the setting. Therefore, the following practice was taken into consideration while preparing the raters for the large-scale assessment of the single paper given to each of them.

- White (1984) puts forward some practice to maintain high reliability in large-scale assessment:
- At least two independent raters must score each script, and if there is any discrepancy a third rater's score may be taken into consideration.
- Scoring should be done in a controlled reading, that is to say, a group of readers meets together to grade scripts at the same place and time. There are two advantages of doing so: (a) the circumstances under which scripts are read are controlled. Therefore there will be a low variance as unnecessary sources causing erroneous scoring will be eliminated, and (b) a positive social environment which helps to set the rating standards is formed.
- Checks on the reading in progress by reading leaders (also called *Table Leaders*) help to ensure that every reader participated in the process is maintaining the agreed-upon standards for grading.

Evaluation and record keeping are essential for the ongoing assessment program. So unreliable raters may be eliminated (cited in Weigle, 2002; p129).

Results

The distributions of two groups of raters' holistic and analytic marks given to a single paper were shown in the below tables. The scores were analyzed by the statistical software SPSS 16.00 v.

Table 4.1. The Comparison of Native Speakers' Holistic and Analytic Scores

Native Speakers' Markings	Participants							
	1	2	3	4	5	6	7	8
Holistic Marks	13	12	13	13	12	13	13	12
Analytic Marks	16	13	14	16	13	14	16	14

Table 4.1. shows the marks given by 8 native English teachers. It can be easily realized that the marks given by analytic scoring is regularly a small amount higher than holistic scoring.

Table 4.2. The Comparison of Non-Native Speakers' Holistic and Analytic Scores

Non-Native Speakers' Markings	Participants							
	1	2	3	4	5	6	7	8
Holistic Marks	15	11	11	16	13	15	11	11
Analytic Marks	12	13	14	16	12	12	13	14

In table 4.2., the marks given by 8 non-native speaker teachers of English depending on analytic and holistic scales were slightly different from those of native ones. It can be realized easily that the marks given by analytic scoring were to some extent higher than holistic scoring (the mean of holistic marks is 12.8750; the mean of analytic marks is 13.2500).

The distribution of the marks given by native speakers (NS) and non-native speakers (NNS) can be seen on the table 4.3 and table 4.4.

Table 4.3. The Marks Given by the Native Speakers

Participant	Holistic Scores	Analytic Marks							Total Score
		Content	Organization	Cohesion	Vocabulary	Grammar	Punctuation	Spelling	
1	13	2	2	2	3	2	2	3	16
2	12	2	2	2	2	2	1	2	13
3	13	2	2	1	3	2	2	2	14
4	13	2	2	2	3	2	2	3	16
5	12	2	2	2	2	2	1	2	13
6	13	2	2	1	3	2	2	2	14
7	13	2	2	2	3	2	2	3	16
8	12	2	2	2	2	2	2	2	14

Table 4.4. The Marks Given by the Non-Native Speakers

Participants	Holistic Scores	Analytic Marks							Total Score
		Content	Organization	Cohesion	Vocabulary	Grammar	Punctuation	Spelling	
1	15	1	2	1	2	2	2	2	12
2	11	1	2	2	2	2	2	2	13
3	11	2	2	2	2	2	2	2	14
4	16	3	3	2	2	2	2	2	16
5	13	2	1	3	1	2	1	2	12
6	15	1	2	1	2	2	2	2	12
7	11	1	2	2	2	2	2	2	13
8	11	2	2	2	2	2	2	2	14

The mean of holistic marks scored by NS raters was a little lower than analytic ones. Yet, it can be said that there was no significant difference between two types of scores (see table 4.5).

Table 4.5. Paired Sample Statistics: Native Speakers' (NS) Scoring Holistic and Analytic Marks

NSs' Scoring Marks	N	Minimum	Maximum	Range	Mean	Std. deviation	Std. Error Mean
Holistic	8	12,00	13,00	1,00	12,6250	,51755	,18298
Analytic	8	13,00	16,00	3,00	14,5000	1,30931	,46291

Table 4.6. shows that NNS gave more points in their analytic ratings than holistic ratings. And also this situation is similar to the results shown on Table 4.5.

Table 4.6. Paired Sample Statistics: Non-Native Speakers' (NNS) Scoring Holistic and Analytic Marks

NSs' Scoring Marks	N	Minimum	Maximum	Range	Mean	Std. deviation	Std. Error Mean
Holistic	8	11,00	16,00	5,00	12,8750	2,16712	,76619
Analytic	8	12,00	16,00	4,00	13,2500	1,38873	,49099

Table 4.7. and table 4.8. show that there are high correlations between both groups raters' holistic and analytic marks ($p > 0,05$ level).

Table 4.7. Paired Sample Correlations: Native Speakers' Scoring Holistic and Analytic Marks

Pair		N	Correlation	Significance
1	Holistic and Analytic Marks	8	,738	,037

Table 4.8. Paired Sample Correlations: Non-Native Speakers' Scoring Holistic and Analytic Marks

Pair		N	Correlation	Significance
2	Holistic and Analytic Marks	8	,059	,889

Discussion and Conclusion

Assessing L2 writing, in fact, is a subjective process where there is no exact correct answer that could be given to any test item when compared to other types of testing language skills such as multiple choice tests. However, to obtain high level of objectivity in assessing writing there are some requirements that scorers must fulfil. Therefore, the term 'inter-rater reliability' gains importance in scoring marks to L2 written productions. In this study, the amount of consistency between the ratings of both the groups of native speakers and non-native speakers were statistically analyzed. Scoring marks depending on analytic and holistic scales showed that both groups of raters are consistent in their analytic and holistic ratings. The correlation of native speakers' scoring holistic and analytic marks is ,738 and significance is ,037 at $p > 0,05$ level and the correlation of non-native speakers' scoring holistic and analytic marks is ,059 and significance is ,889 at $p > 0,05$ level. However, the mean of native speakers' analytic scoring marks is a little higher than their holistic scoring marks. Similarly, the mean of non-native speakers' analytic scoring marks is a little higher than holistic ones. Overall, native and non-native raters rated the same paper consistently (12-15= good). In conclusion, the study is limited to 16 participant-raters and to replicate for confirming the outcomes of the present study, the participants and task may be varied in the further studies.

References

- American Psychological Association *Standards for Educational and Psychological Testing* (1985). Washington, D.C: American Psychological Association.
- Bachman, L. F. (1995). *Fundamental Considerations in Language Testing*. Oxford: Oxford university Press.
- Barkaoui, K. (2010). Variability in ESL essay rating process: The role of the rating scale and rater experience. *Language Assessment Quarterly*, 7, 54-74.
- Bresciani, M. J., Oakleaf, M., Kolhorst, F., Nebeker, C., & et al. (2009) Examining design and inter-rater reliability of a rubric measuring research quality across multiple disciplines. *Practical Assessment Research & Evaluation*, 14, 12, p.3.
- Brown, G.T.L., Glasswell, K., & Harland, D. (2004). Accuracy in the scoring of writing: studies of reliability and validity using a new zealand writing assesment system. *Assessing Writing*, 9, 106.
- Coombe, C., Folse, K., & Hubley, N. (2007). *A Practical Guide to Assessing English Language Learners*. US: The University of Michigan Press.
- East, M. (2009). Evaluating the reliability of a detailed analytic scoring rubric for foreign language writing. *Assessing Writing*, 14, p.91.
- Hamp-Lyons, L. (1990). Second language writing: assessment issues. In Coombe, C., Folse, K., Hubley, N. (2007). *A Practical Guide to Assessing English Language Learners* (p.84). US: The University of Michigan Press.
- Hyland, K. (2003). *Second Language Writing*. Cambridge: Cambridge University Press.
- Stemler, S.E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating, interrater reliability. *Practical Assessment Research & Evaluation*, 9, 4.
- Sudweeks, R. R., Reeve, S., & Bradshaw, W. S. (2005). A comparison of generalizability theory and Many-Facet Rasch Measurement in an analysis of college sophomore writing". *Assessing Writing*, 9, 239.
- Sweedler-Brown, C. O. (1985). The influence of training and experience on holistic essay evaluation. *English Journal*, 74, 5, 49-55.
- TEEP *Test of English for Educational Purposes*, Associated Examining Board, UK, 1984. Weir, C. J, (1990). *Communicative Language Testing*. NJ: Prentice Hall Regents.
- UCLES International Examinations in English as a Foreign Language General Handbook, 1987.
- Weigle, S. C. (2002). *Assessing Writing*, Cambridge: Cambridge University Press.
- White, E.M. (1984). Holisticism. *College Composition and Communication*, 35, 4, 400-409.
- White, E.M. (1985). *Teaching and Assessing Writing*. San Francisco, CA: JosseyBass.

Appendix A. Holistic and analytic scales used in the study.

Figure 1: Holistic Scale, **Source:** UCLES International Examinations in English as a Foreign Language General Handbook, 1987

Grade	Characteristics
Excellent 18-20	Natural English with minimal errors and complete realization of the task set.
Very Good 16-17	More than a collection of simple sentences, with good vocabulary and structures. Some non-basic errors.
Good 12-15	Simple but accurate realization of the task set with sufficient naturalness of English and not many errors.
Pass 8-11	Reasonably Correct but awkward and non-communicating or fair and natural treatment of subject, with some serious errors.
Weak 5-7	Original vocabulary and grammar both inadequate to the subject.
Very Poor 0-4	Incoherent. Errors show lack of basic knowledge of English.

A. Relevance and Adequacy of content

- 0. The answer bears almost no relation to the task set. Totally inadequate answer.
- 1. Answer of limited relevance to the task set. Possibly major gaps in treatment of topic and/or pointless repetition.
- 2. For the most part answers the task set, though there may be some gaps or redundant information.
- 3. Relevant and adequate answer to the task set.

B. Compositional Organization

- 0. No apparent organization of content.
- 1. Very little organization of content. Underlying structures not sufficiently apparent.
- 2. Some organization skills in evidence but not adequately controlled.
- 3. Overall shape and internal pattern clear. Organization skills adequately controlled.

C. Cohesion

- 0. Cohesion almost totally absent. Writing is so fragmentary that comprehension of the intended communication is virtually impossible.
- 1. Unsatisfactory cohesion may cause difficulty in comprehension of most of the intended communication.
- 2. For the most part satisfactory cohesion though occasional deficiencies may mean that certain parts of communication are not always effective,
- 3. Some use of cohesion resulting in effective communication.

D. Adequacy of Vocabulary for Purpose

- 0. Vocabulary inadequate even for the most basic parts of the intended communication.
- 1. Frequent inadequacies in vocabulary for the task. Perhaps frequent lexical inappropriacies and/or repetitions.

- 2. Some inappropriacies in vocabulary for the task. Perhaps some lexical inappropriacies and/or circumlocution.
- 3. Almost no inappropriacies in vocabulary for the task. Only rare inappropriacies and/or circumlocution.

E. Grammar

- 0. Almost all grammatical patterns inaccurate.
- 1. Frequent grammatical inaccuracies.
- 2. Some grammatical inaccuracies.
- 3. Almost no grammatical inaccuracies.

F. Mechanical Accuracy I (Punctuation)

- 0. Ignorance of conventions of punctuation.
- 1. Low standard of accuracy of punctuation.
- 2. Some inaccuracies of punctuation.
- 3. Almost no inaccuracies of punctuation.

G. Mechanical Accuracy II (Spelling)

- 0. Almost all spelling inaccurate.
- 1. Low standard of accuracy in spelling.
- 2. Some inaccuracies in spelling.
- 3. Almost no inaccuracies in spelling.

Figure 2. Analytic Scale, Source: *Test of English for Educational Purposes*, Associated Examining Board, UK, 1984.

Appendix B. Sample section from the composition scored by 8 NS and NNS raters.

= EDUCATIONAL PROBLEMS =

The education problems ~~find~~ in many children, teenagers and adults. The ~~ways~~ which they find difficult and the way these difficulties can be addressed are often different to the approaches commonly applied by schools and are specific to them even though there has been a great deal of work on learning styles over the last two decades. For example, Dum and dum focused on identifying relevant stimuli that may influence learning and manipulating the school environment, at about the same time as Josep Perulli recommended varying teaching strategies. On the other hand, Howard Gardner identified individual talents or aptitudes in his multiple intelligences theories. Many educationalists are against to this system because it gets difficult to ~~teach~~. Although there are these kinds of systems, students don't study some lessons very well. Especially while students learning English as a professional in English departments, some lessons are difficult for them such as, writing, reading and speaking.

Firstly, students produce characteristic, recurring writing problems as a result of inappropriately applying visual thinking to writing. This stems from generally lack of words, unfamiliarity with the kind of analysis and difficulty understanding that context. Some thinkers are in favour of reading books. The more they read books, the more their vocabulary will develop. According to some educationalists, it is not enough to reading books, it also needs to develop their knowledge about the purposes and forms of writing, basic writing skills strategies for planning and evaluating their work and motivation.

Another problem relates to reading, because of reading problems, students do not understand easily some reading passages. First, appropriate levels of reading texts should be chosen for students. They should be slightly above student's English proficiency. Krashen and Terrell argue that input should be at slightly above the current level of proficiency and comprehensible. On the other side, according to Kitao and Kitao, readings should not include too many unknown words or much jargon or dialect or too many technical terms and idioms. Also, they suggest that the content