



# Collaborative Writing Assessment: A Forum for Critical Practice

Masoume Ahmadi and Naser Sabourian Zadeh  
Islamic Azad University, Shahrekord Branch, Iran.

### ARTICLE INFO

#### Article history:

Received: 24 November 2013;

Received in revised form:  
23 December 2013;

Accepted: 6 January 2014;

#### Keywords

Critical thinking;  
Collaborative assessments;  
Individual assessment,  
Think aloud protocol.

### ABSTRACT

In this information era in which people are facing huge amounts of incoming information and communicating too many people who are trying to convince others, teachers' ability to think and to teach critically gains an ever greater saliency as a prime goal of education. To this end, the present study was an attempt to investigate whether collaborative assessment can enhance teachers' critical thinking abilities through empowering them. The collected data consisted of think aloud protocols of 24 individual and 12 collaborative raters. The results were qualitatively analyzed in search of themes relevant to *empowerment* as a critical educational practice among raters/teachers. Transcription of think aloud protocols, comments written down by raters, marginal notes of papers, and qualitative analyses revealed that collaborative assessment as an activity in EFL pedagogy seems to empower raters and provide them with opportunities to express their 'voice'. The findings of this study provide insight into a more effective assessment which contribute to enhancing teachers/raters' critical thinking and present suggestions for further research.

© 2013 Elixir All rights reserved

### Introduction

As is well known, education is of paramount importance for development of society and individual growth of people living in that society. The more developed and civilized one country is, the more well-educated people there are and vice versa. Although societies are developing at a surprising speed, the content and manner of our teaching, somehow, are not. Ian (2007) has described this issue as a common phenomenon in which theories are distinguished from practice and some materials in many textbooks are out of date. He has also mentioned John-Dewey's arguing that "too static aims and materials is opposed acquaintance with a changing world" (p. 57). Accordingly, it can be inferred that there is still a long way to be taken to achieve our educational ideals. A great deal of attention in learning and pedagogy should be given to the sociohistorical and political forces residing in both the meanings of the linguistic resources and the social identities of language users, as Hall (1997) and Bartolome (2004) have argued. Regarding the manner of teaching, Quian (2007) explained that teachers' role as transmitters of knowledge has been taken for granted and has proclaimed them as absolute authorities in the classrooms and students are not supposed to challenge such authorities and foster opinions and thoughts. Teachers as both victims and practitioners of this long term educational practice find it burdensome to face novel ideas and viewpoints of students and even themselves. Regarding the undeniable role of teachers in classrooms, it goes without saying that, teachers should be considered as one of the key factors in determining the success of education and more specifically language teaching. Accordingly, it is not surprising if a broad range of language research endeavors has addressed the characteristic features of successful language teachers. Regarding teaching as an interactive process between society and classroom (Dheram, 2007), considerable debate has taken place over enhancing teachers and learners' critical thinking abilities in EFL/ESL context in recent years. Unfortunately, in this attempt, the teaching community has not been successful in evolving

indigenous critical tools without which empowerment remains only a far-reaching dream.

Nearly everyone agrees that critical thinking has begun to play an outstanding role in education and turned to one of the main goals of education. Changes in technology and workplace have made the ability to think critically more pivotal than any time before. This ability to be involved in an ongoing reflection process is a prominent characteristic of an educated person. Ennis (1992) has mentioned that his concept of critical thinking encompasses seeking a clear statement of the thesis or question, seeking reasons, trying to be well-informed, taking into account the total situation, keeping in mind the original and/or basic concern, being open-minded, and so on. In similar vein, to Stout (1993), Critical thinking has been viewed as an intellectually disciplined process of actively and skillfully conceptualizing, applying, analyzing, synthesizing, and/or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning, or communication, as a guide to belief and action. Critical thinking is "purposeful, reasoned, and goal oriented" (Halpern, 1999, p. 70). It refers to cognitive skills which can be employed in order to solve problems, formulate inferences, calculate likelihoods, and make minds. Although the priorities in today's classrooms include learners' critical thinking abilities, little attention has been paid to this skill from the side of teachers as practitioners and mentors of these abilities in such classes. Recent research suggests that critical thinking is not typically an intrinsic part of tasks at any level. Accordingly, learners' critical thinking should not be considered as an automatic by-product of teaching as it is often taken for granted. A number of researchers (e.g. Bataineh & Zghoul, 2006; Brown, 1984; Hayes & Alvermann, 1986; Kumaravadevelu, 2006; to name a few) claim that the classroom environment must provide modeling, rehearsal, and coaching for students and teachers alike to develop a capacity for informed judgments, for without critical thinking systematically integrated into instruction, learning is transitory and superficial.

Tele:

E-mail addresses: [nasersabourianzadeh@gmail.com](mailto:nasersabourianzadeh@gmail.com)

© 2014 Elixir All rights reserved

Numerous findings (Canagarajah, 2005; Ghahremani Ghajar & Mirhosseini, 2005; Haller, Child, & Walberg, 1988; Reinertsen & Wells, 1993; Stout, 1993; Tsui, 1999; Whitney, 2002; Zohar & Dori, 2003) in the field of education have led a number of professional organizations to stand for highlighting the significance of critical thinking in classroom and to urge teachers to assist their learners in developing higher level thinking skills. Accordingly learners' classroom has seemed to be the logical place to initiate and teach critical thinking. In this regard, Fairclough (1995) has claimed that "if problems of language and power are to be seriously tackled, they will be tackled by the people who are directly involved", and their success depends upon the "theoretical and analytical resources" they have access to (p.221). They require "programmes of ...critical language awareness... to develop the capacities of people for language critique, including their capacities for reflexive analysis of the educational process itself."(ibid., p. 221). Birjandi and Bagherkazemi (2010) found out a statistically significant relationship between Iranian EFL teachers' critical thinking ability and their professional success. They indicated that 'drawing inferences', 'interpreting evidence', and 'evaluating arguments'—as three of five aspects of critical thinking— positively related to the scores obtained from teachers' professional success. Taking a close look at the literature, critical thinking can permanently be considered as an unfinished project (for more information, See Facione, 1990; Gent, 1996; Halvorsen, 2009; Kuhn, 1999; Maleki & Habibipour, 2007; Mason, 2007; McPeck, 1981; Pennycook, 1990; Ramasamy, 2011; Tassel-Baska, 2009) , a story looking for an ending that will never arrive. Accordingly, looking new techniques which move learners and their teachers more forward toward a continual, ongoing search for better opinions, decisions, or judgments will be worthy to be invested. Teachers need to investigate the best methods and techniques in order to integrate them into their classrooms and model good critical thinking practices to their learners. In this journey, teachers' own critical thinking abilities will certainly be of essence due to the fact that it is teachers' duty to prepare their students to function in an increasingly complex society. Choy and Oo (2012) have taken a look at this issue from another viewpoint and mentioned another facet in improving and applying critical thinking in classrooms. They have claimed that reflective thinking as a precursor for incorporating critical thinking is not deeply reflected in teachers' teaching practices. Their findings revealed that critical thinking is practiced minimally.

Many techniques such as journal keeping, writing, reading comprehension have been introduced thanks to their effect on enhancing critical thinking abilities. Learners' Collaborative learning has been put as a subcategory of these techniques too. Reinertsen and Wells (1993) suggested that many students can achieve some awareness of the analytical perspectives and can learn self evaluation, in part, through the use of team journals. Gokhale (1995) and Totten, Sills, Digby and Russ (1991) advocated the role of collaborative learning in enhancing pupils' critical thinking. The result of the study carried out by Quitadamo, Brahler, and Crouch (2009) revealed that peer-led team learning has a small but positive impact on critical thinking gains in some science courses which consequently result in improving grade performance and retention. McConnell (2002) elaborated on the effect of collaborative assessment in e-learning among learners. The results revealed that collaborative assessment increases their self-awareness. Students develop skill and know about self and peer assessment and see themselves as

competent in making judgments about their own and each other's work, which are surely good lifelong learning skills. In spite of these advantages, most of research studies on collaborative works have been done on learners. At yet, there is little empirical evidence on its effectiveness for teachers. This study examined the effectiveness of individual assessment versus collaborative assessment in enhancing teachers' critical thinking skills.

Collaborative assessment (Somervell, 1993) as an alternative form of assessment and contrary to traditional assessment is represented in following diagram (Figure 1, diagrams 2–5). As it can be observed, 5 types of assessment are presented. The first diagram is assessing in traditional form in which teacher is an omnipotent and source of knowledge. Learners submit their writing papers to their teachers and teachers assign scores to the text based on their own criteria. The second diagram shows the form of assessment we have addressed in the present study. In this form the shared activity typically undertaken by two colleague assessors. The third and the fourth types of assessment are peer-assessment and self-assessment respectively. The last diagram represents another form of collaborative assessment and its difference with the second type lies in the act of collaboration which takes place not by two college assessor; in this form a teacher and a student collaborate interactively.

Two major theoretical learning theories support collaborative notion. The first one which arises from the work of Long (1996) is the psycholinguistic theory of interaction. Based on "interaction hypothesis", providing comprehensible output and negative feedback during an interaction facilitate learning. 'Sociocultural theory' is the second of which originates from the work of Vygotsky (1978). This theory highlights the importance of collaboration as well as social interaction. It claims that a human being cannot be developed cognitively unless he/she interacts with others in a society. In its original sense such interactions should be in form of 'scaffolding' (Wood, Burner, and Ross, 1976). It means a novice member of society must rely on a more able individual to reach a higher level of cognitive development. The importance of collaboration has also been emphasized by Messic (1975) who claimed that "researchers, other educators and policy makers must work together to develop means of evaluating educational effectiveness that accurately represent a school or district's progress toward a broad range of important educational goals". The research Literature reveals that a positive and social climate is of paramount importance in developing and sustaining collaborative assessment. Regarding writing assessment, collaborative assessment recedes the traces of the subjectivity and bias of writing assessment and provides a relatively agreed-upon-criteria on which learners can rely after long and short discussions.

To bridge the mentioned gap—lack of research on enhancing teachers' critical thinking – the current research has taken over the mission to find an answer to the following question:

- Is there any significant difference between teachers' critical thinking abilities before and after being involved in individual and collaborative assessments?

### **Methodology**

Qualitative and quantitative procedures were used in the collection and analysis of the data in the study. Because little is known about the content of raters' reflection, individually or collaboratively, the study is exploratory and descriptive in nature to arrive at the basic information.

### Participants

As many as 24 Iranian male and female adult teachers of English as a foreign language (EFL) participated voluntarily in this study. The reason for using volunteer participants was to make sure that they would participate willingly because writing assessment is a difficult and time-consuming job for most teachers and they do not enjoy it. Another reason was to make sure that they would participate in both phases, that is, individual and collaborative assessment. Some researchers (Brown, 1991; Mendelsohn and Cumming, 1987) have figured out that raters from different disciplines apply different criteria to non-native English writing texts. Based on this finding, the final pool of participants were nonnative English teachers majored in TEFL. Consequently, all participants were MA students studying TEFL at Allameh Tabataba'i University, Tehran. The participants aged between 23 and 28 with the median age of 25.8. The raters were different in terms of level of teaching, ranging from elementary to advance. The years of experience in English language teaching ranged from 1 year to 7 years; therefore, participants were novice and professional nonnative language teachers.

### Instruments

Sample writing papers used in this study included 3 parts. The first part contained some personal questions on the characteristics of the participants such as gender, years of experience, level of teaching, etc. The second part composed of 3 texts written by IELTS (GT) examinees. High-stakes tests such as IELTS have a significant impact on the lives of individuals or on programs, and are not easily reversed, so that errors in these decisions can be difficult to correct. Being informed of the nature of this test, the raters showed much more sensitivity about correcting and scoring writing samples. The third part was blank space to be filled with score and scoring criteria by rater. These papers were word processed (not hand written) in order to remove any effect that handwriting could have.

### Procedure

In order to fulfill the objective of the present study, certain procedures were undertaken. First, raters were given instructions which clarified what they were supposed to do such as: scoring IELTS writing samples, providing comments on texts, scoring individually and collaboratively, and thinking aloud while assessing texts. They also were informed that candidates were supposed to write their essays in minimum 250 words. In the next step, writing samples were distributed among raters. The sample writings used were anonymous. Context variables such as time and day of rating session, and ordering of compositions deserve more attention due to the fact that next to the ability of the writers themselves, the largest source of variance came from the raters. Researchers kept these variables fairly constant and all the raters assessed the texts before their class time at university in order to make sure that effect of tiredness and other affective factors is eliminated.

Raters, at first, scored the writing individually while verbalizing their thoughts during scoring. Moreover, they provided the researchers with some comments on writing papers. Participated in this step, they were asked to score collaboratively a blank copy of the same writing samples in paired groups. In doing so, contrary to restrictions of IELTS exam, scoring rubric was provided neither for individuals nor for collaborative pairs; conversely, were they requested not to base their assessments on any of these scales, but to construct their own criteria individually or collaboratively. Erdosy (2001)

has explained that this condition is imposed on raters in order to extract their preferences in scoring and the criteria for assigning scores but not to justify an existing rating scale.

In furnishing concurrent verbal protocols, raters were instructed to think loudly, speak continuously in Persian or English without any time limit set for the task. The introspective verbal protocols were recorded and subsequently transcribed by researchers. The collected data were coded in the search of empowerment words by researchers and, to insure inter-coder reliability, by two other researchers. The basic principle guiding the coding of rater's think aloud data was capturing the kinds of empowerment words and ideas heeded by raters while thinking aloud. The coded data provided a source of information to be analyzed. Chi-Square analyses were conducted to investigate significant differences between individual and collaborative groups. The content analysis of think aloud protocols provided further support.

### Coding the Data

Data was analyzed using a procedure of data reduction and confirmation of findings. The taped sounds were carefully transcribed and coded. The coding categories for the present study were developed based on Toulmin's (1985) because Brown and Keeley (1994) counted these abilities as components of critical thinking. This model is what Stapleton (2001) has described as a logical way for measuring critical thinking. Many studies (Bataineh & Zghoul, 2006; Ennis, 1993; Stapleton, 2001; among others) have been conducted to investigate assessing students' critical thinking abilities through using various well-known tests such as Cornell Critical Thinking Test, California Critical Thinking Skills Test, and Watson-Glaser Critical Thinking Appraisal. There are many other tests but, as Davidson and Dunham (1997) have explained, they are mostly limited to multiple-choice instruments that do not allow learners to provide the reasons behind their answers. Transcribed texts were finally analyzed for the following components:

a. *Number of arguments*: providing arguments would either means being "agree" or "disagree" with the passage but such definitive conclusion is not of essence (Stapleton, 2001). To emphasize on arguing but not drawing any conclusion, Stapleton added that "One can even remain undetermined and undecided and be a good critical thinker" (ibid, p. 516). These arguments concluded or followed by reasons can be presented by claim markers such as "I think" and "in My opinion" and assertions made by some modals such as "should" and "must" which all are followed by phrases such as "because", "due to", and "for this reason". If there is no reason, then the argument is flawed.

b. *Extent of evidence*: Evidences take many forms, including personal experience, research studies, statistics, citing authorities, analogies, pointing out consequences, and precisely defining words. (Stapleton, 2001)

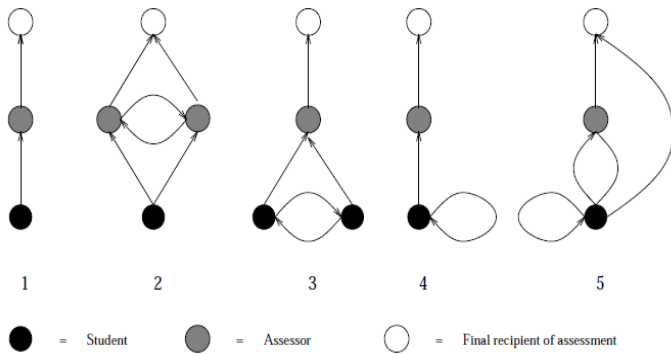
c. *Recognition of opposition*: This subcategory includes recognizing opposite views in order to refuse or challenge them via the presence of specific structures such as "Although it is said ....., ....." .

d. *Corresponding refutations*: It refers to refusing a claim or assumption which may seem incorrect. Refutations must be logically related to the opposing views which they profess to counter. The same as previous category they will be recognized through conjunctive devices such as 'although', 'however', 'even though', etc.

e. *Number of fallacies*: Fallacies are different types of errors in reasoning. Ramage and Bean (1999, cited from Stapleton, 2001) have classified fallacies into three types. The first one has been

called fallacy of pathos which characterizes flaws in the relationship between argumentation and audience. Ethos fallacies are characterized as flaws in the relationship between argument and the character of those involved. The third type, logos fallacies, is known as the relationship between claim and reasons or evidence in argumentation.

Figure 1. Forms of Assessment (Somervell, 1993)



The decisions made upon analysis were a single or several utterances with a single aspect of the event as the focus. Such units may contain just one clause or many clauses focusing on a dominant component of critical thinking. Stapleton (2001) demonstrated these components in more details which were of great aid for the current study too (See Appendix). For further elaboration, consider the following examples:

**Example:**

**Rater 8:** *The major supports should be mentioned here again.* (Argument, Individual assessment)

**Rater 9:** *Man fek mikonam k in neveshte aslan thesis statement nadare, motivator dare ama blue print ham nadare. Badam ink b nazare man transfer va translation touye in matn kheyli ziyade dar vaghe az first language translate karde.*

**English:** *I think, although motivators have been used in this text, no blue print or thesis statement is observable. By the way, in my idea, we can see many signs of transfer and translation here. In fact he has translated from L1.* (Argument, Individual assessment)

**Collaborative raters 10:** *-B nazare man y moshkel k dashtesh in bood k kheyli ba ink nemidounam L1 chi boode kheyli translation dasht toosh.*

*-are kamelan movafegham. English nabood kheyli az sakhtarash.*

**English:** *In my idea, one of the problems of this text is translation from L1.*

*-I do agree, most of sentences were not based on English structure.* (Argument, Collaborative assessment)

**Rater 6:** *The writer is an Iranian person, not authentic and natural.* (Fallacy due to oversimplification, Individual assessment)

**Rater 22:** *First of all, no sensible person can deny the importance of breaking down barriers between countries. Who says that? We may have sensible persons who can break down the barriers between countries. Well there are barriers among countries.* (Refutation, Individual assessment)

**Rater 22:** *if we go one step a head of syntax and go to the semantic, the primary and secondary school they are actually ....it is not one of the important subjects. They have more important stuff to study. History is just fun and they do not have that. It is not real history. It is just stories.*

**Collaborative Raters 11:** *It is not of concern here in this text, but there is a lack of understanding of the topic.* (Recognition of opposition, Collaborative assessment)

**Rater 13:** *Farsi structure is used in this text. Bringing "in" is a kind of Farsi structure. In Farsi we say 'dar pishrafte'....* (Evidence, Individual assessment)

Researchers independently coded transcribed texts for all critical thinking components. In order to improve inter-coder reliability, texts were coded by 2 other researchers. There were discrepancies on how to classify such responses. Therefore, any disagreement about coding was discussed among raters to raise the level of agreement.

**Data analysis**

To probe any significant difference between raters' critical thinking in collaborative and individual groups, Chi-square analyses were run using the 16<sup>th</sup> version of the Statistical Package for Social Sciences (SPSS). The results were supported by further explanation.

**Results**

This study was an attempt to investigate whether significant differences exist between teachers' critical thinking abilities when rating individually and collaboratively. To answer this question, the total number of each CT component was separately counted and Chi-square analyses were run to probe significant differences (Table 1).

Total number of assertions which were based on critical thinking components was equal to (n=226) statements for individual raters out of a 216-minute transcription and (n=241) statements out of a 79-minute transcription in collaborative group. Although no significant discrepancies appeared in Table 1 regarding each individual components except refutations ( $\chi^2=11.7 > \chi^2_{critical} = 3.84$ ), the total amount of Chi-square observed value ( $\chi^2=33.02, df=4, p=.000 \leq .05$ ) is much more higher than the critical value of 9.48 at 4 degrees of freedom. Therefore, it can be concluded that there are significant differences among critical thinking components of both groups of raters.

**Table 1. Analysis of Chi-square on the Use of Critical Thinking Components between Collaborative and Individual Assessments**

Section		Collaborative and Individual Assessment
Argument	Chi-Square	.004
	df	1
	Asymp. Sig.	.000
Evidence	Chi-Square	.93
	Df	1
	Asymp. Sig.	.000
Recognition of Opposition	Chi-Square	3.35
	Df	1
	Asymp. Sig.	.000
Refutation	Chi-Square	11.77
	Df	1
	Asymp. Sig.	.000
Fallacy	Chi-Square	.96
	Df	1
	Asymp. Sig.	.000
Total	Chi-Square	33.01
	Df	4
	Asymp. Sig.	.000

As it is displayed in Table 2, collaborative raters exceeded the expected rate in employing some components of critical thinking such as 'arguments' and 'refutations'. They have used 'arguments' 0.8 times more than expected and 'refutations' as

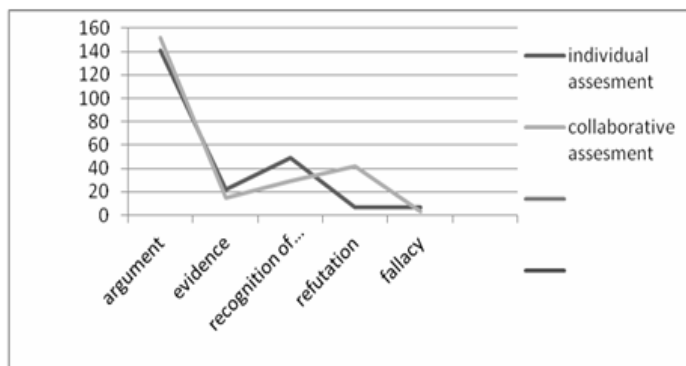
much as 16.7 times while their selection of ‘evidence’, ‘recognition of opposition’, and ‘fallacy’ choices are lower than the expected rate. Interestingly, they have recognized the opposite views about 11 times more than expected rate and consequently, ‘recognition of opposition’ was placed as the lowest component of critical thinking used by collaborative raters.

**Table 2. Observed, Expected and Residuals in collaborative assessments**

		Observed N	Expected N	Residual
<b>Collaborative assessment</b>	Argument	152	151.2	+8
	Evidence	15	19.09	-4.09
	Recognition of opposition	29	40.25	-11.25
	Refutation	42	25.28	+16.72
	Fallacy	3	5.1	-2.1

Figure 2 presents the difference between using each component of critical thinking in individual and collaborative groups. Collaborative raters outperformed individual group through more arguing and refusing. Regarding providing evidence, recognizing oppositions, and making fallacies, the graph shows a downward trend.

**Figure 2. The Use of Critical Thinking Components between Individual and Collaborative Raters**



## Discussion

Over the past decades, studies have continued to suggest the significance of critical thinking and reasoning. “The ability to think critically is needed in this revolutionary age of technology change. Among the essential skills required to close the gap between the knowledge and skills people learn in schools and those required to function effectively in the workplace and community is the ability to think critically” (Lee, 2007, p. 82). Recent studies (e. g. Dinkelman, 2000; Gómez, 2010; Lightbown, 2000; Okazaki, 2005) have also emphasized the highly influential role of teachers in creating a classroom environment that promotes critical thinking and claimed that in order to make this possible, EFL teachers should be allowed pre-service and in-service training opportunities to encourage the development of critical thinking. Critical thinking skills are vital to make the best possible choices and using resources to the greatest advantage. This study also was an attempt to find out whether one form of alternative assessment, namely collaborative assessment, is influential in fostering critical thinking abilities. The analyses revealed that raters become more critical while rating collaboratively rather than individually. Some other researchers (Gokhale, 1995; McConnell, 2002; Quitadamo, et al., 2009; Totten, et al., 1991) second the findings. Gokhale (1995) has elaborated on the role of collaborative learning in enhancing pupils’ critical thinking. Arend (2009) has explained that higher levels of critical thinking were related to discussions among learners and the way these

discussions are carried out. Tsui (1998, 1999) has also accentuated the role of discussion on improving critical thinking abilities and skills. Online discussions typically have the purpose of creating a space and time for informal, open-ended thinking to occur. Critical thinking appears to be best encouraged among students when a more consistent emphasis is placed on the discussions.

The findings of this study presented insights into how raters operate across individual or collaborative assessment. The major findings emerging from these analyses reveal that how enhancement of critical thinking abilities in collaborative assessment will facilitate raters with undeniable benefits. Before starting with how to test writing, we must start with attempting to define what we mean by writing ability. This is a big deal since as, researchers in L1 and L2 have pointed out, the uses of which writing is put out by different people in different situations therefore, and no single definition can cover all situations (Camp, 1993; Purves, 1992). Raters may apply not necessarily the same criteria to the scripts, and even when they apply the same criteria, their evaluations and judgments may differ. Teachers bring their teaching experience to the task of judging and engage themselves as active readers trying to make common sense with students’ writings. Linacre (1989) counted rater variation as an undesirable problem that causes error variance and this problem should be eliminated or reduced, but there are some problems in achieving this goal. Linacre (1989) believes that the phenomenon of rater variation is an inevitable part of rating process. Investigating rating criteria has attracted the attention of many other precedent researchers (see for example Cumming, 1989; Cumming, 2002; Eckes, 2005; Erdosy, 2001; Siyyari, 2011; Son, 2010; Wang, 2010; Weigle, 1994). As Cumming, Kantor, and Powers (2001) pointed out, there are number of criteria which fall into the categories that are pertinent as performance criteria is increasing due to the common experience of language teaching and testing. The variability due to rater characteristics which has been called *rater effect* and involves many consequential impacts on decision-making processes has been investigated by some other researchers (Barrette (2001), Myford and Wolf (2000), and Hsieh (2011), among others). Matsumoto and Kumamoto’s (n.d) study revealed that the inter-related variables noticeably influence the emphasis place in the evaluation of writing. Such variables include nationality, the type of training the raters had received, their teaching experience, and their approaches and tenets as a teacher. Using Jacobs, et al (1981), they also concluded that greater differences caused by aforementioned factors were outstanding in the categories of content, grammar, and organization.

Messick (1998, cited in Hsieh, 2011) has stated that these effects undermine the validity and fairness of performance assessments. Mei (n.d.) declared that it is essential for raters to remain objective as far as possible to allow their students to develop their positions. One of the advantages of collaborative assessment is that raters move away from their biases and subjective point of view. McConnell (1999) corroborated this idea and claimed that collaborative assessment strives to bring about a variety of viewpoints and values to the assessment process and in doing so make the process of assessment more accountable and open. “The openness of collaborative assessment process is crucial to its success” (McConnell, 2002, p. 89). This seems to be confirmed through the findings of the current study.

Elder and Paul (2010) have elaborated on the issue that critical mind consciously seeks the truth in accordance with the following maxims:

- “I believe it, but it may not be true.”
- “We believe it, but we may be wrong.”
- “We want to believe it, but we may be prejudiced by our desires or cultural limitations.”
- “It serves our vested interest to believe it, but our vested interest has nothing to do with the ethical reality.” (P.37)

Accordingly, a critical thinker will decrease or even eliminate the impact of raters' effects as much as possible by acting upon those aforementioned maxims. Ennis (1993) has also reiterated the characteristic of critical thinkers. He believes that critical thinkers need to judge the credibility of sources, identify conclusions and reasons as well as assumptions, judge the quality of an argument, including the acceptability of its reasons, assumptions, and evidence, develop and defend a position on an issue, be open-minded, try to be well-informed and draw conclusions when warranted, but with caution. This critical thinker, as Halpern, 1999 mentioned, uses these skills appropriately, without prompting, and usually with conscious intent, in a variety of settings. Based on these claims, it is crystal clear that collaborative assessment can render assessment to a less subjective one through enhancing critical thinking.

Moreover, it can be observed that collaborative raters have the lower rate of using fallacies comparing with individual ones; however, the difference is not eye-catching. Fallacies are erroneous thinking which play trick on one's critical thinking. Ramasamy (2011) has mentioned that informal fallacy as one of the dimensions of critical thinking can distract people in thinking critically because “they tend to appear reasonable and their unreliability is not apparent on the surface” (p. 1). Based on this claim, the finding of this study on enhancing critical thinking while assessing collaboratively can be approved of. A good critical thinker has the ability of recognizing these fallacies and reasoning them out before deciding or concluding certainly.

### Conclusion

On the one hand concerns about reliability and validity in assessment have been voiced and, on the other, critical thinking as a never-ending story needs to be researched continuously. Critical thinking is more than just knowledge acquisition or a collection of processing skills; rather it is the development and continual use of analytical skills (Scriven & Paul, 2005). The analyses in this study have looked for a technique of assessment through which critical thinking enhance. Such an influence will consequently decrease the negative impact of rater effect. The present findings are hoped to have significant implication(s) for EFL/ESL teachers in general and Iranian EFL teachers in particular. At this point, a word regarding limitations of this study is worth mentioning. The findings of this study should be taken cautiously due to gathering of information just through think aloud protocols. Observations, diaries, retrospective process, and interviews can help next researchers to triangulate the data. Further comparisons could reveal more about the relationship between critical thinking skills and other prominent pedagogical solutions in ELT.

### References

Arend, B. (2009). Encouraging Critical Thinking in Online Threaded Discussions. *The Journal of Educators Online*, 6(1), 1–23.

Batainneh, R. F. & Zghoul, L. H. (2006). Jordanian TEFL graduate students' use of critical thinking skills (as Measured by

the Cornell Critical Thinking Test, Level Z). *The International Journal of Bilingual Education and Bilingualism*, 9 (1), 31–49.

Birjandi, P. & Bagherkazemi, M. (2010). The Relationship between Iranian EFL teachers' critical thinking ability and their professional success. *English Language Teaching*, 3(2).

Brown, I. D. (1991). Do English and ESL learners' faculties rate writing samples differently? *TESOL QUARTERLY*, 25(4), 587–603.

Bartolomé, L. I. (2004). Critical pedagogy and teacher education: Radicalizing prospective teachers. *Teacher Education Quarterly*, 97–122.

Brown, M. N. & Keelley, S. M. (1994). *Asking the right questions*. Englewood Cliffs, NJ: Prentice Hall.

Canagarajah, S. (2005). Critical pedagogy in L2 learning and teaching. In E. Hinkel (Ed.), *Handbook on research in second language teaching and learning* (pp. 931-950). Mahwah, NJ: Lawrence Erlbaum.

Cumming, A. Kantor, R. & Powers, D. E. (2001). *Scoring TOEFL essays and TOEFL 2000 prototype writing Tasks: An investigation into raters' decision making and development of a preliminary analytic framework*. TOEFL Monograph series, NJ: ETS.

Cumming, A. (1998) Theoretical perspectives on writing. *Annual Review of Applied Linguistics* 18, 61–78.

Cumming, A. (2002). Assessing L2 writing: Alternative Constructs and ethical dilemmas. *Assessing Writing*, 8, 73–83.

Davidson, B. W. & Dunham, R. A. (1997). Assessing EFL student progress in critical thinking with the Ennis–Weir critical thinking essay test. *JALT Journal*, 19, 43–57.

Dinkelman, T. (2000). An inquiry into the development of critical reflection in secondary student teachers. *Teaching and Teacher Education*, 16, 195–222.

Elmore, P. B. & Farley, M. J. (1992). The relationship of reading comprehension to critical thinking skills, cognitive ability, and vocabulary for a sample of underachieving college freshmen. *Educational and Psychological Measurement*, 52(4), 921–931.

Ennis, R. (1993). Critical thinking assessment. *Theory into Practice*, 32(3), 179–187.

Erdosy, M.U. (2001). The influence of prior experience on the construction of scoring criteria for ESL compositions: A case study. *IJES*, 1(2), 175-196.

Garrison, D. R. (2011). Critical thinking and self-directed learning in adult education: An analysis of responsibility and control issues. *Adult Education Quarterly*, 42(3), 136–148.

Gokhale, A. A. (1995). Collaborative learning enhances critical thinking. *Journal of Technology Education*, 7(1), 22–30.

Gómez, S. P. (2010). Building students' voices through critical pedagogy: Braiding paths towards the other. *Colomb. Appl. Linguist. Journal*, 12(2), 55–71.

Hall, G.S. (1997). *Redefining the Syllabus: An investigation into whether syllabuses can meet learners' linguistic and social needs'*. Essay for CCLL1 Module, MA ELT, Lancaster University.

Haller, E.P., Child, D.A. and Walberg, H.J. (1988). Can comprehension be taught? A quantitative synthesis of 'metacognitive' studies. *Educational Researcher* 17 (9), 5–8.

Halpern, D. H. (1999). Teaching for critical thinking: Helping college students develop the skills and dispositions of a critical thinker. *New Direction for Teaching and Learning*, 80, 69–75.

Hsieh, C. N. (2011). Rater effects in ITA testing: ESL teachers' versus American Undergraduates' judgments of accentedness,

comprehensibility, and oral Proficiency. *Spaan Fellow Working papers in Second or Foreign Language Assessment*, 9, 47–74.

Lee, K. (2007). Online collaborative case study learning. *Journal of College Reading and Learning*, 37(2), 82-100.

Lightbown, P. M. (2000). Anniversary article: Classroom SLA research and second language teaching. *Applied Linguistics*, 21, 431-462.

McConnell, D. (1999). Examining a collaborative assessment process in networked lifelong learning. *Journal of Computer Assisted Learning*, 15, 232–243.

Mei, W. S. (n.d). Investigating raters' use of analytic descriptors in assessing writing. *Reflections on English Language Teaching*, 9(2), 69–104.

Mirhosseini, S. A. & Ghajar, S. G. (2005). English class or speaking about everything class? Dialogue journal writing as a critical EFL literacy practice in an Iranian high school. *Language, Culture and Curriculum*, 18 (3), 286-300.

Quitadamo, J., Brahler, J., & Crouch, G. J. (2009). Peer-Led Team Learning: A Prospective Method for Increasing Critical Thinking in Undergraduate Science Courses. *Science Educator*, 18 (1), 29–40.

Qian, L. (2007). Critical thinking and educational ideal. *US-China Education Review*, 4(1), 174- 202.

Reinertsen, S. P. & Wells, C. M. (1993). Dialogue journals and critical thinking. *Teaching Sociology*, 21 (2), 182-186.

Seid Motahhari, M. (1998). Critical thinking as ideology dialogic in construction chaotic in action. An unpublished MA thesis. Allameh Tabataba'i University, Tehran, Iran.

Stapleton, P. (2001). Assessing critical thinking in the writing of Japanese university students: Insights about assumptions and content familiarity. *Written Communication*, 18, 506-550.

Stout, C. J. (1993). The dialogue journal: a forum for critical consideration. *Studies in Art Education*, 35(1), 34-44.

Siyari, M. (2011). On the cross relationship between self/peer assessment: Rating accuracy and personality traits, and the contribution to writing performance. Unpublished Doctoral dissertation, Allameh Tabataba'i University, Tehran, Iran.

Somervell, H. (1993). Issues in assessment, enterprise and higher education: The case for self-, peer and collaborative assessment. *Assessment and Evaluation in Higher Education*, 18(3), 221–233.

Son, M. (2010). Examining rater bias: An evaluation of possible factors influencing elicited imitation ratings. Unpublished Master Thesis, Young University, Brigham.

Toulmin, S. E. (1958). *The use of argument*. Cambridge, UK: Cambridge University Press.

Tsui, L. (1999). Courses and instruction affecting critical thinking. *Research in Higher Education*, 40 (2), 185–200.

Wang, B. (2010). On rater agreement and rater training. *English Language Teaching*, 3(1), 108-112.

Weigle, S. C. (1994). Effects of training on raters of ESL compositions. *Language Testing*, 11, 197–223.

Zohar, A. and Dori, Y.J. (2003) Higher order thinking skills and low-achieving students: Are they mutually exclusive? *The Journal of the Learning Sciences*, 12 (2), 145-182.

## Appendix

### Rubrics for elements of critical thinking

#### Conclusion

A conclusion is a statement or series of statements in which a writer sets out what she wants the reader to believe. This belief

is conveyed via the argument, evidence, and other statements that the author uses to signal his belief. Conclusions are usually limited to agreeing, disagreeing, or taking some middle ground with respect to the prompt. Conclusions may be explicit or self-announcing, using clear language. Such conclusions are often preceded by declarations such as “I agree”, “I disagree”, “I am undecided”, or indicator words or phrases including “therefore”, “instead”, and “as a result”. Conclusions may also be implicit, or unfolding, where the belief is unstated and left to infer.

#### Argument

Each argument consists of a claim supported by a reason. A claim consists of a statement whose truth is arguable, and is often advanced in answer to a problem or controversial issue. A claim which stands alone without a supporting reason is an opinion and cannot be classified as an argument. Claims are often expressed using claim markers such as “I think” or “in my opinion”. Claim assertions can come in various forms including proposals, evaluations, and definitions.

Reasons are statements used to support claims and generally answer why the claim should be believed. Reasons must show a direct logical link to the claim in order to be bound into a single proposition called an argument. Reasons need to be new; however, if they are a simple repetition without elaboration, they do not indicate critical thinking. Reasons are often identified by indicator words and phrases such as “because”, “for this reason”, and “for one thing”.

#### Evidence

Evidence constitutes statements or assertions which serve to strengthen the argument. Evidence comes in many forms including: personal experience, research studies, statistics, citing authorities, comparison and analogies, pointing out consequences, facts, logical explanation, precisely defining words.

#### Recognition of Opposition and Refutation

Opposing viewpoints constitute statements that run counter or offer alternative interpretations to those expressed in the claim. Refutations are statements in which the response is given to the opposing viewpoint in a way that shows that is inadequate in some way. Ramage and Bean (1999) explained that shortcomings in opposing viewpoints can include logical flaws, poor support, erroneous assumptions, or wrong values. Refutations must be logically linked to the opposing views which they profess to counter. They can also offer rival causes or solutions. In refuting an opposing or alternate view, the writer maintains his conclusion. Opposing viewpoints and refutations are identified by indicator phrases and words such as,

a. “it is said that....but,)”

b. “some people claim that...however,”

c. Conjunctive devices, including “although”, “despite” and “even though”

#### Fallacies

Fallacies are errors in reasoning. They occur when the reason does not adequately support the claim in one of a number of ways. They include oversimplification, irrelevant assertions, etc.