

The Effect of Instructing Critical Thinking through Debate on Male and Female EFL Learners' Reading Comprehension

Maryam Danaye Tous¹, Abdorreza Tahriri², and Sara Haghghi³

Abstract: The purpose of the present study was to examine the effect of instruction through debate on male and female EFL learners' reading comprehension. Also, their perception of critical thinking (CT) instruction was investigated. A quantitative research method with experimental pre-and post-tests design was conducted to collect the data. Eighty-eight EFL learners- who were selected via convenience sampling method- were randomly assigned to two experimental groups (22 males and 22 females) and two control groups (22 males and 22 females). Data were analyzed using descriptive and inferential statistics. The Oxford Placement Test (OPT) was administered to choose the intermediate sample. To ensure the homogeneity of the participants in terms of reading skills, the Reading Comprehension Placement Test (RCPT) was conducted. Also, the California Critical Thinking Skills Test (CCTST) and Read Theory Critical Reading Comprehension Test (RTCRCCT) were used as pre-and post-tests to measure the students' CT skills. Although the findings showed that debate had a statistically significant effect on EFL learners' reading comprehension ability, the role of gender was not found to be significant. In addition, the results revealed that there was no significant difference between male and female EFL learners' perception of CT instruction. It was concluded that instructing CT skills through debate resulted in a better understanding of the reading texts.

Keywords: reading comprehension, critical thinking, debate technique, gender.

Introduction.

In recent decades, studies on reading comprehension have led to great emphasis on the important role of problem-solving techniques that supposedly enable the students to identify, evaluate, and solve perplexities that arise in reading (Waters, 2000). According to Stancato (2000), researchers agree that creativity, problem-solving, and imagination of one's comprehension processes are critically important aspects of skilful reading. Such imagination and creativity are often referred to in the literature as critical thinking (CT) (Stancato, 2000). Facione and Facione (1994) also stated that CT is the process of analysis, evaluation, inference, deductive reasoning, and inductive reasoning.

Using analysis, one can express and comprehend the significance of a wide variety of experiences, data, beliefs, conventions, and criteria (Facione & Facione, 2010). Using evaluation, one can decide how weak or strong an argument may be, and the credibility of statements or descriptions of a person's perception, judgment, or opinion could be assessed (Facione & Facione, 2010). Using inference, one can identify elements needed to draw reasonable conclusions based on evidence and reason to form hypotheses. Also, consequences from

¹ Assistant professor of linguistics, University of Guilan, Rasht, Iran, Maryam.dana@gmail.com

² Assistant professor, PhD in TEFL, University of Guilan, Rasht, Iran, atahriri@gmail.com

³ M.A in TEFL, University of Guilan, University Campus 2, Rasht, Iran, sara.pardis@gmail.com

opinions, principles, beliefs, questions, or other forms of representation could be deduced (Facione & Facione, 2010). Using deductive reasoning, one can determine if a conclusion is true or if the premises leading to it are true (Facione & Facione, 2010). Moreover, using inductive reasoning, one can generalize from specific pieces of evidence to valid results and conclusions (California Academic Press, 2006).

Stapleton (2001) claimed that CT is an important factor in the acquisition of reading. Also, Osborne (2005) believed that in order to demonstrate the ability to read critically, debate is an effective technique. According to Freeley and Steinberg (2005), CT that includes debate allows for collaboration where teams can achieve higher levels of thinking through the use of persuasive evidence. This collaboration allows individuals to retain information longer and provides them with an opportunity to engage in the discussion and shared learning (Freeley & Steinberg, 2005). Freeley and Steinberg (2005) define debate as the process of advocacy and inquiry, a way of arriving at a reasoned judgement on a proposition. Snider and Schnurer (2002) also mentioned that in-class debate cultivates the active engagement of students. Thus, the students' approach changes from a passive learner to an active one.

Whereas the debate technique requires all students to actively engage in the multidimensional teaching and learning of a topic area, the lecture format allows them to receive and respond to instruction (Omelicheva & Avdeyeva, 2008). Roy and Macchiette (2005) stated that debate techniques are better suited for the enhancement of CT skills than traditional techniques such as lecture. Studies comparing lectures versus debates found that those students who were exposed to debates perform better on comprehension tasks (Omelicheva & Avdeyeva, 2008). Because the Meeting-House Debate strategy was used in this study, it will be explained here. In this strategy, each side gives its opening argument, and then the rest of the class question the debaters or offer comments. Also, the teacher, acting as a moderator, ensures that each team receives questions equally. Finally, each side gives its final argument (Chial & Riall, 1994).

Furthermore, a few studies have examined the effect of gender on the CT skills. For example, Walsh (1996) found females to be superior to males at higher order thinking, whereas, traditional beliefs and stereotypes claimed that men are superior at analytical thinking (cited in Barjesteh & Vaseghi, 2012). In the present study, gender is considered as one of the independent variables relevant to the CT skills.

Literature Review.

Historical Background.

The term CT dates back to 2500 years ago. Socrates laid the first foundation for analytical examination of basic assumption, determining cause and effect of speech and action, and finding evidence (Cosgrove, 2009). Further, debate as a teaching strategy dates back to over 2400 years to Protagoras in Athens, the father of debate (Freeley & Steinberg, 2005).

Studies on CT and Reading Comprehension.

Studies using quantitative methods report some benefits of CT skills for English as a foreign language (EFL) learners' reading comprehension. For example, Barjesteh and Vaseghi (2012) carried out a study to investigate the possible effect of CT training on EFL learners' reading comprehension. The participants were divided into two low and high proficiency groups and

each group was further divided into critical and non-critical groups. The results of their study confirmed the effect of CT training on the learners' reading comprehension. Also, Aloqaili (2011) examined the correlation between CT and reading comprehension. The results of this study revealed that there was a well-established relationship between CT and reading comprehension. In another study, Fahim, Bagherkazemi, and Alemi (2010) explored the relationship between the performance on the reading section of the paper-based TOEFL and the CT skills. Three tests, including WGCTA-Form A, the reading section of the paper-based TOEFL, and the reading section of general training IELTS were administered. The results of their study indicated that there was a positive relationship between the two variables. The relationship between CT skills and reading comprehension was also tested by Sheikhy Behdani (2009) and Lachini (2003) who came up with a meaningful relationship between these two variables.

Study on CT and Debate.

Goodwin (2003) was among the first researchers who studied the students' perception toward the debate technique. In this study, all the students worked in teams to prepare debates on issues arising from reading and lecture. The groups presented debates, and those not debating acted as judges and wrote a brief essay expressing their views. Some students reported that the new technique was uncomfortable. However, a lot of students expressed that the debate technique was very helpful in gaining knowledge and helped them with analyzing arguments. Also, they believed that debate helped them keep an open mind to the opinions of others and it improved their CT skills.

Studies on CT and Gender.

A few studies found a significant relationship between CT skills and gender. On the one hand, the findings of Walsh (1996) revealed that females had higher levels of CT skills than those of males. On the other hand, the results of a study by King, Mines, and Wood (1990) showed that CT scores of graduate students differed by gender. In their study, males scored higher than females. In another study conducted by Claytor (1997), gender was found to be independent of CT skills.

Studies on Reading Comprehension, CT, and Debate.

Regarding Iranian studies, Rashtchi and Sadraeimanesh (2011) investigated the effect of using debate strategy on EFL learners' reading comprehension. Two homogeneous groups of 55 students were randomly assigned to the control and experimental groups. In the experimental group, the debate strategy was used whereas the control group followed the traditional reading procedures. Findings revealed that the debate strategy had a significant effect on reading comprehension. In addition, Fahim and Saeepour (2011) investigated the effect of instructing CT skills on reading comprehension ability, and the effect of using debate strategy on CT skills. Sixty intermediate students were assigned to the control and experimental groups. The students who represented the experimental group received some treatment using debate format. Findings revealed that the difference between the control and experimental groups' performance on the

CT test was not significant, but the difference between them in terms of reading comprehension performance was significant.

Only a few studies have examined the effect of instructing CT skills through debate on EFL learners' reading comprehension. Hashemi (2011) stated that Iranian educational system emphasizes transmitting information and limits students' learning to memorizing the materials. In other words, the majority of students in Iranian EFL context are not educated as thoughtful individuals (Fahim & Saeepour, 2011). Thus, the big problem facing Iranian EFL learners is that when they are given the reading materials which are ambiguous to them, they cannot disambiguate confusion and think through the problem.

This study sought to investigate the effect of instructing CT skills through debate on male and female EFL learners' reading comprehension, and also to examine the difference between them in terms of their perception of CT instruction. To this end, the following research questions were posed:

1. Does instruction through debate have any significant effect on male and female EFL learners' reading comprehension?
2. Is there any significant difference between male and female EFL learners' perception of CT instruction?

Methods.

Research Design.

This study was done using a quantitative research method with two designs: experimental pre-post tests and a quantitative content analysis design, respectively. *Independent variable:* The first independent variable (instructional technique) varied over two levels, the instructional technique implemented in the experimental group using the Meeting-House Debate strategy and the traditional technique using the lecturing strategy implemented in the control group. The second variable was student gender (male vs. female). A third independent variable (participant) varied over two levels, the control and experimental groups. *Dependent variables:* The dependent variables were the students' pre-and post-test scores on the Read Theory Critical Reading Comprehension Test (RTCRCCT) and California Critical Thinking Skills Test (CCTST).

Participants.

The research population included 120 high school male and female students (11th graders), in Lahijan City located in Guilan Province, Iran. Out of 120 students, 88 of them including 44 males and 44 females- who had three to five years' experience of private English classes- were selected as the research sample, based on the convenience sampling method. Then, they were grouped into the control (22 males and 22 females) and experimental groups (22 males and 22 females). It should be noted that a statistical power analysis was run based on data from a pilot study. The effect size in this study was 1.0, which could be considered to be large, using Cohen's (1988) criteria. With an alpha = 0.05 and power = 0.80, the sample size needed for this between/within group comparison with this effect size was N=60. Thus, the sample size of 88 was adequate for the purpose of this study.

Materials.

The reading materials were selected from New Interchange series (Richards, 2007) including *The Truth About Lying*, *The Global Village*, *A Day in Your Life-In the year 2020*, and *Are You in Love?* .In addition to these required reading texts, a separate handout on a controversial topic such as *Love and Lie* taken from Wikipedia was distributed among students in the experimental and control groups.

Instruments.

The instruments involved in this study consisted of the Oxford Placement Test (OPT), Reading Comprehension Placement Test (RCPT), RTCRCT, CCTST, and a questionnaire (see appendix). In order to ensure the homogeneity of the participants as intermediate learners, the OPT consisting of 50-items was administered.

Furthermore, two types of reading tests, including the RCPT and RTCRCT were used. Based on the results of the RCPT as the pre-test, a homogeneous group was selected. The reason behind administering the RTCRCT was the importance of comprehension in terms of CT skills. It was administered to all participants in the control and experimental groups as pre-and post-tests to measure deductive reasoning, conclusion making, logical inference, sequential analysis, total awareness, and understanding of scope.

The first reading test, i.e., RCPT, was designed so that students could take two tests. The first one (Test 1) was a screening test that required written responses and was administered to the entire class. Students who committed more than seven errors on the screening test took a second test (Test 1.1) that placed them in Comprehension A group. Students who committed seven or fewer errors on the screening test took another test (Test 1.2) that placed them in Comprehension B group. The screening test (Test 1) was made up of 16 multiple-choice items. Students were asked to complete it in 10 minutes. Test 1.1 contained 18 items, and it took around 10 minutes. Test 1.2 was a written test containing four items. Students underlined sentence parts, wrote answers to questions, and indicated correct responses to multiple-choice items. This test required 10 minutes to be completed.

The second reading test, i.e., RTCRCT was a literal reading comprehension test included three passages followed by 24 multiple-choice items. Although this test was designed to prompt the students to think critically, it extracted their CT skills implicitly. This is the difference between this test and the following test, i.e., CCTST, which explicitly measures five dimensions of CT (i.e., analysis, evaluation, inference, deductive reasoning, and inductive reasoning).

The CCTST consists of two Forms (A and B). Items of Forms A and B are parallel according to responses and questions. They contain 34 multiple-choice items of different levels of difficulty and can be administered in a 45-minute period. In addition, the CCTST is composed of a total CT skills score. The total score is considered to be a valuable predictor of success for the completion of educational programs, licensure examinations, and certification.

Above all, a separate questionnaire was used to study the perceptions of the experimental groups toward CT instruction. The questionnaire consisted of 30 multiple-choice items developed by Fahim and Saeepour (2011) and four open-ended questions added by the present researchers. The permission to use this questionnaire was obtained from its authors. In order to achieve a better understanding of the clarity of items and instructions, a pilot study was conducted with 24 students (11th graders), 12 high school males and 12 high school females. According to the participants' comments in the pilot study, three out of seven expository

questions which addressed the same issues were deleted. In addition, one complex question was broken down into two simple ones.

Reliability and Validity.

To ensure the reliability, the OPT was piloted on a sample- who was selected randomly- consisting of 15 males and 15 females of 11-graders. In this study, Cronbach alpha coefficient was found to be 0.80. The RCPT was piloted on a group of 12 male and 12 female students. They were selected randomly from the 11th graders who were studying at the same high school of the main participants. They were asked to answer the same reading test. The reliability of the test was calculated using Cronbach alpha value ($r = 0.79$).

Furthermore, the RTCRCT was piloted on a group of 10 male and 10 female students. They were selected randomly from the 11th graders. They were asked to answer the same reading test as the main test. In this study, Cronbach alpha coefficient was 0.77.

Regarding the Persian version of the CCTST- Form B, Cronbach alpha coefficient for the reliability was 0.71. Depending on the testing context, KR-20 alphas range from 0.70 to 0.75 (Facione, Facione, & Giancarlo, 2000). The confidence coefficient is 0.62 and the construct validity is between 0.60-0.65 with highly positive correlation (Khalili & Soleimani, 2003). In addition, the reliability of the Persian version of the questionnaire was measured via Cronbach alpha ($r = 0.81$). Its face and content validity was confirmed by two experts in the field (University assistant professors, Ph.D. in TEFL). In the current study, the construct validity of the questionnaire was examined using exploratory factor analysis.

Procedure.

For the purpose of this study, the experimental group was randomly assigned to two debater groups, and 12 students known as “debriefers” who were responsible for asking questions, offering comments, giving critical opinions, and asking for reason. Also, each debater group was made up of five members. All students in each group of debaters were expected to work together, and try to persuade the debriefers to accept their perspectives. They were asked to be prepared for the possible arguments against them. To this end, they were asked to surf on the internet, make use of any other available sources, and get extra information required to defend their opinions. During the debate sessions, students known as debaters and debriefers always had the same role.

The debate sessions were conducted on controversial topics that would lead to the debaters’ disagreement. At the end of the debate sessions, the debriefers were asked to pose questions, ask for clarification, and examine the closing arguments within the timeframes (e.g. 10 minutes). Then, they were asked to vote and report back to the debaters. The debriefers’ judgments encouraged the debaters to present the arguments based on relevant or real cases. However, presenting the relevant instances was not always in hand. Therefore, the debaters could not often justify their perspectives or support reasons with a good example. This would lead the debriefers to show the different opinions or give critical views.

Prior to the beginning of the debate session, all the debaters and debriefers were given a brief explanation of the debate etiquette. They were told that all the students would be responsible for their comments. Further, they were asked to focus attention on refraining from saying you are

wrong, attacking the idea and not the person, avoiding exaggeration, avoiding quarrelling, and watching their tone of voice.

To choose a topic that was interesting for the students, the instructor used *brainstorming* strategy. First, a table of various topics was drawn. The list of topics was determined by the participants in both experimental and control groups. Thus, the subject matters could not account for some of the gains in either group. Three topics were chosen through a simple voting process. All the students grouped as teams were asked to write on the board a topic or an area of subject that interested them. They could choose as many topics as they wanted. They were then told that they had only one vote. As a result, the total number of votes for each subject was calculated. The most popular topics developed by the students are presented in the following table.

Table 1. Debatable topics

Olympic	Customs	Actors	Technologies	Entertainments	General
Games	Thanksgiving	Biography	Robots	Computer games	Love
Stars	Halloween	Works	Cars	Facebook	Lie
Referees	Mother's day	Payments	Computers	Sports	working on weekends
Events	Valentine's day	Possessions	Cell phones	Music	Smoking

After this step, the students were taught how to ask someone for his/her opinions, how to interrupt, how to ask for information, etc. For example, they were asked to interrupt with “May I add something?” .Also, they were taught a few examples of widely-used expressions like (a) agreeing: That’s exactly what I think; (b) disagreeing: I don’t think so!; and (c) irony expressions: Are you kidding?

The debate sessions were conducted for different lengths of time: 20-30 minutes; 25-35 minutes; and 30 to 45 minutes. In the classroom, one-piece seats for three students were fixed because there was no other way to arrange the desks. Therefore, all the students sat in rows and no specific shape like “U” was used. Also, the speaking time was divided equally between the two debating teams.

The debate started with the affirmative team and it was followed by the opposing team. At first, a member of the affirmative team presented his/her argument. Then, the speaker on the opposing team presented his/her opposing argument. Next, further arguments supporting the previous arguments were presented by one of the affirmative speakers. After that, one of the opposing speakers identified further areas of conflict, attempting to argument against it and defending his/her opposing argument. Finally, the debate teams received varied feedback about their performance from the researcher and the debriefers.

At the end of the debate sessions, the debriefers were asked to evaluate each debate team individually. A list of criteria for assessment of the debaters’ performance was developed by the researchers. The list consisted of specific aspects of quality such as knowledge on topics, use of examples, use of gestures for clarity, speaking in a clear-cut way, persuasive presentation, strong arguments, ability to present counter-argument, and drawing conclusions. Furthermore, the rubric designed by Glantz and Gorman (1997) was used to get a better understanding of students’ performance. The rubric consisted of: (a) Is the student well organized? (b) Does the student focus on the central ideas of the debate? (c) Is every statement supported by cited researched evidence? (d) Is the research recent? (e) Is the research complete? (f) Is an adequate number of sources used? (g) Is the evidence presented with bias in some way? (h) Does the student make frequent eye contact with the audience? (i) Does the student respond to all of the opponents’

points? (j) Does the student challenge flaws in the opposition's arguments? (k) Does the student avoid distorting information, making faulty generalizations, and oversimplifying issues?

Following the debate, the debriefers were also asked to rank their favourite team and choose the group that they found most assertive. In addition, they were asked to write on what they agreed or disagreed about based on the topics discussed in the debate sessions.

It is worth noting that the medium of instruction in the experimental group was English. More importantly, the questions discussed in this group were not in line with what would be given in the post-tests. The experimental and control groups were taught in more or less the same condition except for the treatment. The treatment sessions were held twice a week. After the treatment sessions that lasted for one month and a half, the experimental and control groups took exactly the same post tests. Over the same period of time, the control group received no particular treatment. The participants in the control group received their regular instruction based on traditional technique. According to this technique, students were not required to share ideas, participate in role play, judge beliefs, and engage in discussion. Furthermore, the class was not put into groups.

In order to reduce bias, the subject matters between the debate (experimental) group and the control group was consistent. In the control group, all the students were asked to read the same reading text as the experimental group did. The meaning of unknown vocabularies was given by the teacher. Also, the students could access a dictionary. Then, they were asked to memorize the meaning of new vocabularies. Next, they were required to present a brief summary of the reading texts. It should be noted that all the students in the control group, who were taught through the medium of English, were asked to change speed, avoid back-channel such as "umm", and talk for 8 to 10 minutes. In addition, they were expected to answer the follow-up reading questions. Students' responses were checked, and if incorrect, they were given spoken feedback by the teacher. It should be noted that the questions asked in the control group were not designed to prompt those questions that were included in the post-tests.

Data Analysis.

Both descriptive and inferential statistics were used to analyze the data. To this end, SPSS statistical package, version 20 was used. Measures of central tendency and standard deviation were computed for the pre-and post-test scores. In order to answer the first research question, the data were analyzed using two-way ANOVA. To answer the second research question, the data were analyzed using independent samples t-test to see if there was a significant difference between male and female EFL learners' perception of CT instruction. Besides, students' responses to open-ended questions in the questionnaire were analyzed using the quantitative content analysis method.

Results

Homogeneity of 120 participants in terms of their level of language proficiency was determined by the OPT. Males had a mean of 40.21 with a standard deviation of 5.42. Females had a mean of 38.95 with a standard deviation of 6.22. Forty-eight males scored between 34 and 46 out of 50. Also, Forty-eight females scored between 32 and 46 out of 50. Thus, the participants whose score did not fall within a range of one standard deviation above and below the mean (24 learners) were excluded. Due to the purpose of this study, the homogeneity of selected male and

female participants (96 learners) was also determined in terms of reading skills. The results of RCPT revealed that males had a mean of 22.64 with a standard deviation of 1.99. Females had a mean of 22.97 with a standard deviation of 2.04. As a result, four participants were excluded. Forty-six males and 46 females, who had scored between 20 and 25 out of 35, one standard deviation above and below the mean, were selected.

The selected students (46 males and 46 females) took the RTCRCT that elicits students' CT skills implicitly. Males had a mean of 16.26 with a standard deviation of 1.65. Females had a mean of 16.41 with a standard deviation of 1.69. Forty-four males and 44 females, who had scored between 14 and 18 out of 24, one standard deviation above and below the mean, were selected. In addition, the homogeneity of selected students was assessed via the CCTST that elicits students' CT skills explicitly. Males had a mean of 14.79 with a standard deviation of 2.56. Females had a mean of 14.77 with a standard deviation of 2.50. All of the students were in the score range of 10 to 20 out of 34. No one was excluded. The results showed that 44 males and 44 females could be selected as the main sample.

To ensure that there was no significant difference between the experimental and control groups, and males and females regarding reading comprehension and CT skills, a two-way ANOVA was run, respectively. The results of RTCRCT revealed that the main effect of "participant" was not significant, $F(1, 84) = 0.070, p = 0.895 > 0.05$, and there was no significant main effect for "gender", $F(1, 84) = 2.478, p = 0.792 > 0.05$. Also, the interaction between gender and participant was not significant, $F(1, 84) = 0.158, p = 0.692 > 0.05$. Further, the results of CCTST showed that the main effect of "participant" was not significant, $F(1, 84) = 7.469, p = 0.068 > 0.05$, and there was no significant main effect for "gender", $F(1, 84) = 0.002, p = 0.966 > 0.05$. Also, the interaction between gender and participant was not significant, $F(1, 84) = 0.092, p = 0.762 > 0.05$. Thus, there was not a significant difference between experimental and control groups' scores on the RTCRCT and CCTST. Also, there was no significant difference between males and females' scores at the beginning of this study.

Table 2. Tests of between-subject effects, dependent variable: RTCRCT pre-test scores

Source	Type III Sum of Square	df	Mean Square	F	Sig.
Corrected Model	0.636 ^a	3	0.212	0.082	0.970
Intercept	23563.636	1	23563.636	9090.939	0.000
Gender	0.182	1	0.182	2.478	0.792
Participant	0.045	1	0.045	0.070	0.895
Gender*Participant	0.409	1	0.409	0.158	0.692
Error	217.727	84	2.592		
Total	23782.000	88			
Corrected Total	218.364	87			

a. R Squared = 0.003 (Adjusted R Squared = -0.033)

Tests of between-subject effects, dependent variable: CCTST pre-test scores

Source	Type III Sum of Square	df	Mean Square	F	Sig.
Corrected Model	45.670 ^a	3	15.223	2.521	0.063
Intercept	19234.102	1	19234.102	3185.287	0.000
Gender	0.011	1	0.011	0.002	0.966
Participant	45.102	1	45.102	7.469	0.068
Gender*Participant	0.557	1	0.557	0.092	0.762
Error	507.227	84	6.038		
Total	19787.000	88			

Corrected Total	552.898	87
-----------------	---------	----

a. R Squared = 0.083 (Adjusted R Squared = 0.050)

Table3 presented the differences in mean scores for male and female students in the control and experimental groups on the pre-and post-tests of RTCRCT. This table reports only the descriptive statistics which do not show whether these differences are large enough to be considered statistically significant.

Table 3. Descriptive statistics of males and females’ RTCRCT scores on pre-and post-tests

Scores	N	Participants	Gender	Mean	Std. Deviation
Pre-test	22	Experimental	Female	16.54	1.56
	22	Control		16.36	1.49
	22	Experimental	Male	16.22	1.71
	22	Control		16.40	1.56
Post-test	22	Experimental	Female	20.54	1.50
	22	Control		15.86	1.42
	22	Experimental	Male	19.54	2.30
	22	Control		15.68	1.28

Differences in the mean scores of male and female students on the RTCRCT

Gender	N	Participants	Mean Difference (A)	Std. Deviation
Female	22	Experimental	4.00	1.63
	22	Control	-0.50	1.22
	44	Total	1.75	2.72
Male	22	Experimental	3.32	1.75
	22	Control	-0.72	1.27
	44	Total	1.30	2.54
Total	44	Experimental	3.66	1.71
	44	Control	-0.61	1.24
	88	Total	1.52	2.63

A: the mean score differences are computed through post-test RTCRCT scores subtracted from pre-test RTCRCT scores.

Table4 showed the results of ANOVA for the main effect of gender and participant as two independent variables. It was the pre-and post-test score difference used in running ANOVA. The results revealed that the main effect of “participant” was significant, $F(1, 84) = 184.8, p = 0.000 < 0.05$. This showed that there was a significant difference between experimental and control groups’ scores. The experimental group had a better performance on the post-test. Also, table4 revealed that there was no significant main effect for “gender”, $F(1, 84) = 2.478, p = 0.119 > 0.05$. That is, there was not a significant difference between males and females’ scores on the pre-and post-tests. Further, the interaction between gender and participant was not significant, $F(1, 84) = 0.737, p = 0.393 > 0.05$.

Table 4. Tests of between-subject effects

Source	Type III Sum of Square	df	Mean Square	F	Sig.
Corrected Model	417.364 ^a	3	139.121	62.67	0.000
Intercept	210.182	1	210.182	94.68	0.000
Gender	5.500	1	5.500	2.478	0.119
Participant	410.227	1	410.227	184.8	0.000
Gender*Participant	1.636	1	1.636	0.737	0.393
Error	186.455	84	2.220		
Total	814.000	88			
Corrected Total	603.818	87			

a. R Squared = 0.691 (Adjusted R Squared = 0.680)

Table5 reports the descriptive statistics and the differences in mean scores of males and females on the CCTST.

Table 5. Descriptive statistics of males and females’ CCTST scores on pre-and post-tests

Scores	N	Participants	Gender	Mean	Std. Deviation
Pre-test	22	Experimental	Female	14.13	2.62
	22	Control		15.40	2.26
	22	Experimental	Male	14.00	2.92
	22	Control		15.59	1.89
Post-test	22	Experimental	Female	17.63	3.04
	22	Control		14.59	2.48
	22	Experimental	Male	17.18	3.55
	22	Control		14.54	2.15

Differences in the mean scores of male and female students on the CCTST

Gender	N	Participants	Mean Difference (B)	Std. Deviation
Female	22	Experimental	3.50	1.15
	22	Control	-0.81	1.00
	44	Total	1.34	2.07
Male	22	Experimental	3.18	0.99
	22	Control	-1.05	1.12
	44	Total	1.06	2.04
Total	44	Experimental	3.34	1.07
	44	Control	-0.93	1.06
	88	Total	1.20	2.13

B: the mean score differences are computed through post-test CCTST scores subtracted from pre-test CCTST scores.

To find the probable differences between students’ scores on the CCTST as pre-and post-tests, ANOVA was run. The results revealed that the main effect of “participant” was significant, $F(1, 84) = 217.15, p = 0.000 < 0.05$. This showed that the scores of the experimental group differed significantly from pre-test to post-test to the benefit of post-test. However, there was no significant main effect for “gender”, $F(1, 84) = 1.642, p = 0.204 > 0.05$. This result suggested that male and female students were almost at the same level of CT. Further, the interaction between gender and participant was not significant, $F(1, 84) = 0.026, p = 0.873 > 0.05$.

Table 6. Tests of between-subject effects

Source	Type III Sum of Square	df	Mean Square	F	Sig.
Corrected Model	387.682 ^a	3	129.227	72.942	0.000
Intercept	137.500	1	137.500	77.611	0.000
Participant	384.727	1	384.727	217.15	0.000
Gender	2.909	1	2.909	1.642	0.204
Participant*Gender	0.045	1	0.045	0.026	0.873
Error	148.818	84	1.772		
Total	674.000	88			
Corrected Total	536.500	87			

a. R Squared = 0.723 (Adjusted R Squared = 0.713)

In order to determine the relationship between students’ performance on the RTCRCT and CCTST, Pearson Product-Moment Correlation Coefficient was run. The results showed that there was a positive relationship between variables A and B [$r = 0.723$, $n = 88$, $p = 0.000 < 0.05$]. That is, students who scored higher on the RTCRCT also scored higher on the CCTST.

Table 7. Correlation between A and B

		A	B
A	Pearson Correlation	1	0.723**
	Sig. (2-tailed)		0.000
	N	88	88
B	Pearson Correlation	0.723**	1
	Sig. (2-tailed)	0.000	
	N	88	88

** Correlation is significant at the 0.01 level (2-tailed).

A: the mean score differences are computed through post-test RTCRCT scores subtracted from pre-test RTCRCT scores.

B: the mean score differences are computed through post-test CCTST scores subtracted from pre-test CCTST scores.

Regarding the second research question, the descriptive statistics of students’ responses is presented in table8. Items with a mean score lower than the mid-point (3) indicated the most negative viewpoint, while items with a mean score higher than the mid-point indicated the most positive viewpoint. Thus, Q1 “I make notes on the important elements of people’s argument or propositions” ($M = 1.47$, $SD = 0.50$) gained the most negative viewpoint. One reason could be the limited amount of time in the debate sessions in which students were unlikely to take notes. Also, Q22 “I solicit input from other people to broaden my understanding of a subject” ($M = 4.31$, $SD = 0.80$) gained the most positive viewpoint. In other words, students found questioning peers’ opinions helpful to gain a better understanding of the topics. Further, male and female students displayed considerable positive viewpoints on most of the statements except for Q13 “I play devil’s advocate in order to improve my grasp of an argument or proposition” ($M = 2.68$, $SD = 1.21$). One hypothesis is that participants’ arguments were not strong enough.

Table 8. Descriptive statistics of close-ended questionnaire items

Item	Mean	Std. Deviation	N	Item	Mean	Std. Deviation	N	Item	Mean	Std. Deviation	N
Q1	1.47	0.50	44	Q11	3.88	0.84	44	Q21	3.93	0.89	44
Q2	3.75	0.99	44	Q12	4.13	0.73	44	Q22	4.31	0.80	44
Q3	3.61	1.08	44	Q13	2.68	1.21	44	Q23	3.59	1.08	44
Q4	3.79	1.09	44	Q14	3.59	1.08	44	Q24	3.86	0.90	44
Q5	3.90	0.96	44	Q15	4.04	0.96	44	Q25	3.68	1.02	44
Q6	3.59	0.87	44	Q16	3.31	0.82	44	Q26	3.79	0.97	44
Q7	3.61	1.20	44	Q17	3.70	0.82	44	Q27	3.97	0.87	44
Q8	3.56	1.12	44	Q18	3.75	0.96	44	Q28	4.00	0.83	44
Q9	3.65	1.01	44	Q19	3.72	0.94	44	Q29	3.93	0.92	44
Q10	3.61	0.94	44	Q20	4.11	0.78	44	Q30	3.77	0.80	44

Group Statistics

Gender	N	Mean	Std. Deviation	Std. Error Mean
Female	22	3.70	0.301	0.064
Male	22	3.67	0.243	0.051

The results of t-test for independent samples revealed that there was not a significant difference between male and female students' perception, $t(42) = 0.381, p = 0.705 > 0.05$.

Table 9. Statistical analysis of independent samples t-test of close-ended questionnaire items

Levene's Test for Equality of Variances		Independent samples t-test						
		t-test for Equality of Means						
F	Sig.	t	df	Sig. (2-tailed)	Mean difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
0.889	0.351	0.381	42	0.705	0.031	0.082	-0.135	0.198

Furthermore, all the students' responses to the four expository questions were analyzed using descriptive analysis. The data were transcribed and analyzed for the frequency of positive and negative perceptions. Items were analyzed in order to produce profiles of these students with either positive or negative perceptions. The results of this part are as follows:

Participants in the experimental group were asked if they thought that the debate technique increased their CT abilities. With respect to the first question, nearly all of the students (93.18%) indicated that the debate technique helped them increase their CT skills. They thought that debate provided them with a new opportunity to analyze the data and evaluate the arguments. However, a few students (6.81%) were not satisfied with this technique. They felt that having

students debate a topic did not result in enhancing CT skills. One student said “some of my peers were poor at evaluating”. Regarding the second question, most of the students (88.63%) expressed that they enjoyed working in teams. They indicated that debate was a helpful way of interacting with other students. They agreed that receiving help from their peers during the teamwork was very helpful. In addition, they thought that debate yielded active learning through the process of gathering information and discussing issues. More importantly, some of the students agreed that working in teams helped them gain self-confidence. They claimed responsibility for their learning. Only 11.36% found group work activities less helpful for their learning.

Regarding the third question, most of the students (90.90%) commented on the debate as a good tool for active engagement of students in the classroom. They believed that this technique placed more active role on the shoulders of students. Further, a majority of them reported that getting involved was an important aspect of debate. One student said, “Participating in the debate sessions made everybody involve because it looks like a competitive game”. Another student said, “I have to be ready for oral presentation in the classroom. Also, I have to speak clearly and listen carefully”. The result of this explanatory question provided more support for the result of close-ended items as it showed that students who participated in the debate sessions preferred to listen carefully. On the contrary, 9.09% of the respondents found debate stressful and fatiguing.

The last item addressed the unique value of the debate technique. Some students (11.36%) indicated that they enjoyed speaking in front of the class. Some respondents (13.63%) believed that analyzing arguments and questioning peers’ views helped them gain a better understanding of topics. The students’ explanatory responses supported the result of open-ended questions. In addition, 13.63% of the respondents thought that preparing for debate helped them learn to speak. Some of them (9.09%) thought that friendly atmosphere during the debates allowed them to express their opinions comfortably. They also believed that this technique taught them how to respect their opposite ideas. Moreover, 15.90% of the respondents reported that the debate technique was open to challenge. They found “being challenging” to be the most interesting part of the debate sessions. Some students (13.63%) believed that thinking on the positive and negative sides of topics helped them come up with an appropriate decision. Also, 11.36% of the respondents stated that the debate technique was a new experience.

Table 10. Descriptive statistics of students’ responses

Question	Frequency			Percent		
	Yes	No	No Difference	Yes	No	No Difference
Q1	41	3	–	93.18	6.81	–
Q2	39	5	–	88.63	11.36	–
Q3	40	4	–	90.90	9.09	–
Q4	39	5	–	88.63	11.36	–

Discussion

The first research question addressed the effect of instruction through debate on Iranian male and female EFL learners’ reading comprehension. The increase from pre-test to post-test in the performance of the experimental group (see table3) revealed that the debate technique had a statistically significant effect on the students’ reading comprehension (see table4).That is, the

debate group significantly outperformed the control group. This result is consistent with Rashtchi and Sadraeimanesh's (2011) study, and Fahim and Saeepour's (2011) study in which debate was found to have a significant effect on reading comprehension. The result of this study is also in harmony with that of Barjesteh and Vaseghi (2012) that showed the significant effect of CT training on Iranian EFL learners' reading comprehension. Further, it was revealed that there was a positive correlation between RTCRCT and CCTST scores. This result is in line with that of Fahim, Bagherkazemi, and Alemi (2010) who showed that there was a positive correlation between reading comprehension and CT (see table7).

The results also showed that there was no interaction between gender and participant. It means that there was the same change in the mean scores across gender for both control and experimental groups. Additionally, the statistical analysis of the main effect of gender revealed no significant difference between male and female students. This result did not support the result of Walsh's (1996) study in which females had higher levels of CT skills than males. Similarly, this result did not confirm the findings of King, Mines, and Wood's (1990) study in which males scored higher than females. Reflecting on the findings, gender was shown to be independent of CT skills. This result is in line with that of Claytor (1997) who reported that there was no correlation between gender and CT skills.

Furthermore, the questionnaire responses collected in this study revealed that there was a contradictory belief regarding the debate technique. While most of the students found debate to be useful in developing CT abilities, a few respondents didn't support the idea that debate helped them enhance their CT skills. These findings support the result of Goodwin's (2003) study in which some participants found debate uncomfortable. One of the possible conclusions to be drawn is that some students were not successful in collaborating with their peers. Thus, they failed to spread the feeling of interdependence. Another hypothesis is that they had difficulty in expressing their ideas or defending their opinions.

Conclusion

Traditionally, students are expected to enhance reading comprehension through lecturing method. In this study, the debate technique was used to improve students' reading skills. Considering the limitations (e.g. duration of treatment) and delimitations (e.g. assessing just 11th graders), it was found that improving reading skills through the debate technique was superior to the lecturing strategy. Further, the "Meeting-House Debate" strategy used as the treatment changed the students as passive learners to students as active learners. That is, the debate technique forced the students to learn more broadly through active learning than to learn less broadly through passive learning. Also, the students' responses to the open-ended questions indicated that the majority of the participants had positive views toward the debate technique. The results showed that most of the participants found debate effective and enjoyable. Besides, the correlation between CT and gender was close to zero. It can be concluded that gender did not have an effect on the students' CT skills. Research on the effect of instruction through debate on reading comprehension is not sufficient in many respects. Suggestions for further studies are put forward below:

This study showed that students from the experimental group outperformed the control group. Whether they will be able to transfer what they have been taught to other settings is not clear. Thus, a follow-up study using students selected as the sample in this study is recommended. In

addition, future studies might investigate the effect of instruction through debate on reading comprehension, utilizing random sampling strategy with the larger sample sizes.

Implications of this Study

The findings of this study held important implications for English teachers and material developers. The present study supported the need for teaching CT skills through debate which was shown to be effective for promoting reading comprehension abilities. Instructors' effective use of questions and engaging students in free discussions over controversial and interesting topics could involve students in CT process (Bagherkazemi, Derakhshan, & Rezaei, 2011). Furthermore, findings of this study might encourage the material developers to pay due attention to the key role of CT and debate technique; that is, students' textbooks need to be revised with the aim of enhancing CT skills. Bagherkazemi and Birjandi (2010) also believed that material developers should make an effort to create lessons that promote CT as effective skills connected to academic success.

Acknowledgements

The authors extend their gratitude to Dr. Nader Ofoghi, assistant professor, University of Guilan, for his guidance.

Appendix

Appendix 1. Questionnaire

Dear participants

The following questionnaire is designed for a study on the effect of instructing critical thinking (CT) through debate on male and female EFL learners' reading comprehension. The questionnaire consists of 30 multiple-choice items and 4 open-ended questions. Please read the questions carefully and select one of the options. The choices are:

- Never
- Rarely
- Sometimes
- Often
- Always

Q1. I make notes on the important elements of people's argument or propositions (e.g. topics).

Never Rarely Sometimes Often Always

Q2. I test the assumptions underpinning an argument or proposition.

Never Rarely Sometimes Often Always

Q3. I state my reasons for accepting or rejection arguments and propositions.

Never Rarely Sometimes Often Always

Q4. I put material I have read or seen into my own words to help me understand it.

Never Rarely Sometimes Often Always

Q5. I distinguish between facts and opinions.

Never Rarely Sometimes Often Always

Q6. I double-check facts for accuracy.

- Never Rarely Sometimes Often Always
Q7. I check other people's understanding of issues.
Never Rarely Sometimes Often Always
Q8. I search for parallels and similarities between issues.
Never Rarely Sometimes Often Always
Q9. I use a set of criteria against which to evaluate the strength of the argument.
Never Rarely Sometimes Often Always
Q10. I summarize what I have heard or read to ensure I have understood properly.
Never Rarely Sometimes Often Always
Q11. I breakdown materials so that I can see how ideas are ordered.
Never Rarely Sometimes Often Always
Q12. I asses the credibility of the person presenting the material I am evaluating.
Never Rarely Sometimes Often Always
Q13. I play devil's advocate in order to improve my grasp of an argument or proposition.
Never Rarely Sometimes Often Always
Q14. I set aside emotive language to avoid being swayed by bias or opinionated statement.
Never Rarely Sometimes Often Always
Q15. I evaluate the evidence for an argument or a proposition to see if it is strong enough to warrant belief.
Never Rarely Sometimes Often Always
Q16. I explore statements for ambiguity.
Never Rarely Sometimes Often Always
Q17. I challenge proposals and arguments that appear to lack rigor.
Never Rarely Sometimes Often Always
Q18. I weigh up the reliability of people's opinions.
Never Rarely Sometimes Often Always
Q19. I ask questions to reinforce my understanding of the issue.
Never Rarely Sometimes Often Always
Q20. I establish the assumptions that an argument rests upon.
Never Rarely Sometimes Often Always
Q21. I draw conclusions from data I have analyzed in order to decide whether to accept or reject a propositional argument.
Never Rarely Sometimes Often Always
Q22. I solicit input from other people to broaden my understanding of a subject.
Never Rarely Sometimes Often Always
Q23. I analyze propositions to see if the logic is sound.
Never Rarely Sometimes Often Always
Q24. I set aside my prejudices to evaluate arguments in a dispassionate way.
Never Rarely Sometimes Often Always
Q25. I distinguish major point from minor points.
Never Rarely Sometimes Often Always
Q26. I look for what isn't there rather than concentrate solely on what is there.
Never Rarely Sometimes Often Always
Q27. I reach my own conclusions rather than let myself be swayed by the opinions of others.
Never Rarely Sometimes Often Always

Q28. I research a subject to enhance my understanding.

Never Rarely Sometimes Often Always

Q29. I establish the underlying purpose of an argument or proposition.

Never Rarely Sometimes Often Always

Q30. I consider new information to see whether I need to re-evaluate a previous conclusion.

Never Rarely Sometimes Often Always

1. Do you believe that the debate technique increase your critical thinking abilities? If yes, discuss your point of view.
2. Do group work activities during the debate sessions enhance your learning? If yes, what is your reason?
3. Do you think that the debate technique encourage all students to stay actively engaged in the classroom? If yes, how debate encourages students to take an active role in the classroom?
4. Dose the debate technique have any unique value? If yes, what is the unique value of the debate technique?

References

Aloqaili, A.S. (2011). The relationship between reading comprehension and critical thinking: A theoretical study, *Journal of King Saud University-Languages and Translation*, 24, 35-41.

Bagherkazemi, M., & Birjandi, P. (2010). The relationship between Iranian EFL teachers' critical thinking ability and their professional success. *English Language Teaching*, 3(2), 135-145.

Bagherkazemi, M., Derakhshan, A., & Rezaei, S. (2011). Critical thinking in language education. *Journal of Language Teaching and Research*, 2(4), 769-777.

Barjesteh , H., & Vaseghi, R. (2012). Critical thinking: A reading strategy in developing English reading comprehension performance. *Sheikhbahee EFL Journal*, 1(2), 21-33.

California Academic Press. (2006). *Insight assessment-tools to evaluate reasoning and critical thinking: California critical thinking skills test (CCTST)*. Millbrae, CA: author.

Chial, M., & Riall, A. (1994). *Rules of engagement for classroom, debates*. Retrieved May 9, 2012, from www.comdis.wisc.edu/staff/mrchial/PMT/pdfs/PMTDebates.pdf

Claytor, K.L. (1997). *The development and validation of an adult medical nursing critical thinking instrument (andragogy)*. Doctoral dissertation, Indiana University, Bloomington.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Second Edition. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Cosgrove, R. (2009). Critical thinking in the Oxford Tutorial. Retrieved May 5, 2012, from <http://www.criticalthinking.org/.../d99ff67fa6f1fbcf22084daaaa9efa8951dc7f.com>

Danaye Tous, M., Tahriri, A., and Haghghi, S.

Facione, P.A., & Facione, N.C. (2010). *The California critical thinking skills test manual*. Millbrae, CA: California Academic Press.

Facione, P.A., & Facione, N.C. (1994). *The California critical thinking skills test and national league for nursing accreditation requirement*. Millbrae, CA: California Academic.

Facione, P.A., Facione, N.C., & Giancarlo, C.A. (2000). The disposition toward critical thinking: Its character, measurement, and relationship to critical thinking skill. *Informal Logic*, 20(1), 61-84.

Fahim, M., Bagherkazemi, M., & Alemi, M. (2010). The relationship between test takers' critical thinking ability and their performance on the reading section of TOEFL. *Journal of Language Teaching and Research*, 1(6), 830-837.

Fahim, M., & Saepour, M. (2011). The impact of teaching critical thinking skills on reading comprehension of Iranian EFL learners. *Journal of Language Teaching and Research*, 2(4), 867-874.

Freeley, A., & Steinberg, D. (2005). *Argumentation and debate: Critical thinking for reasoned decision making* (11th ed.). Belmont, CA: Wadsworth.

Glantz, S., & Gorman, B. (1997). He said, she said: Debating with technology. *Technology Connection*, 4(7), 14-16.

Goodwin, J. (2003). Students' perspectives on debate exercises in content area classes. *Communication Education*, 52(2), 157-163.

Hashemi, S.A. (2011). The use of critical thinking in social science textbooks of high school: A field study of Fars province in Iran. *International Journal of Instruction*, 4(1), 64-78.

Khalili, H., & Soleimani, M. (2003). Determining the trust, credibility and norm scores of the California critical thinking skills test-Form B (TSTCC-B). *Journal of Babol University of Medical Sciences*, 2, 84-90.

King, P.M., Mines, R.A., & Wood, P.K. (1990). Critical thinking among college and graduate students. *The Review of Higher Education*, 13(2), 167-186.

Lachini, K. (2003). *Critical thinking and reading comprehension*. Paper presented in the fourth linguistic conference, Allameh Tabatabaie University, Iran.

Omelicheva, M.Y., & Avdeyeva, O. (2008). Teaching with lecture or debate? Testing the effectiveness of traditional versus active learning methods of instruction. *PS*, 41(3), 603-7.

Osborne, A. (2005). Debate and student development in the history classroom. *New Directions for Teaching and Learning*, 103, 39-50.

Danaye Tous, M., Tahriri, A., and Haghighi, S.

Rashtchi, M., & Sadraeimanesh, F. (2011). Is debate a useful strategy in enhancing the reading comprehension and critical thinking of Iranian EFL learners? *Theory and Practice in Language Studies*, 1(4), 361-369.

Richards, J.C. (2007). *New Interchange*. New York, NY: Cambridge University Press.

Roy, A., & Macchiette, B. (2005). Debating the issue: A tool for augmenting critical thinking skills of marketing students. *Journal of Marketing Education*, 27(3), 264-276.

Sheikhy Behdani, R. (2009). *The relationship between autonomy, critical thinking ability, and reading comprehension of Iranian EFL learners*. Master's thesis, Islamic Azad University, Science and Research Campus, Tehran, Iran.

Snider, A., & Schnurer, M. (2002). *Many sides: debate across the curriculum*. New York: International debate education association.

Stancato, F.A. (2000). Tenure, academic freedom, and teaching of critical thinking. *College Student Journal*, 34(3), 377-382.

Stapleton, P. (2001). Assessing critical thinking in the writing of Japanese University students: Insights about assumptions and content familiarity. *Written Communication*, 18(4), 506-548.

Walsh, C. M. (1996). *Critical thinking disposition of university students in practice disciplines (nursing, education, and business) and non-practice disciplines (English, history, and psychology): An exploratory study*. College Park, MD, University of Maryland.

Waters, A. (2000). Thinking and language learning. *English Language Teaching Journal*, 60(4), 237-319.