STUDENTS' CRITICAL THINKING ENHANCEMENT THROUGH SOCRATIC QUESTIONING: A PEDAGOGICAL STUDY

MOOSA AHMED HASSAN BAIT ALI SULAIMAN

Educational Department, College of Arts and Applied Sciences, Dhofar University, Oman
E-mail: moosa@du.edu.om

Abstract - With this attempt as a backdrop that enhancing students critical thinking (CT) in instructional practices is of vital importance in ESL/EFL classrooms, the current study was carried out to investigate the effect of Socratic questioning (SQ) on students CT skills with reference to the five domains of CT 1). Inference, 2). Recognition of Assumptions, 3). Deduction, 4). Interpretation, and 5). Evaluation of Arguments, in grade 12 in Omani female schools. The present study made an attempt to look for differences among students in using CT skills in reading classes. A quasi-experimental design was used in this study. Therefore, a total of 60 EFL female students distributed in two groups. Data analysis showed statistically significant differences between mean scores of those who are taught through SQ and those obtained by students who are taught CT skills in a normal setting. The students in the experimental group managed to implement better CT skills than those who were taught through traditional methods. Also, results showed that CT domains were significantly affected by SQ skills. In view of the discussion of findings pedagogical implications are explored and recommendations made.

Keywords - Evaluative Questions, Probing and Illustrating Questions, Recognition of Assumptions Questions.

I. INTRODUCTION

The Sultanate of Oman is facing the priorities need a high degree of adaptability and a concrete background in school education challenge of educating its youth for life and work in the new priorities made by the current global economy. These particularly in the English language, in order to deal with the modern economy, and capture international business opportunities. Consistent with this understanding, the Sultanate of Oman has specified in the Omani education system that Omani students acquire knowledge and skills in all areas of curriculum including skills in questioning, investigating, critical thinking, problem solving, and decision making (Ministry of Education, 2008). Though it is stated in the national curriculum specifications that "the new English Language curriculum is being designed to equip learners with higher cognitive abilities and skills, and attitudes that Omani learners will need to succeed in this rapidly changing society (Ministry of Education, 2008, p.7)". Teachers' procedures and techniques of questioning may prevent learners from attaining this objective. This is based on the understanding that teacher's knowledge and actual use of questions could influence the way they apply national education programs in their teaching, which in turn may affect the use of classroom questions. This will therefore hamper the nation's effort in achieving its objectives as stated in its educational system.

II. LITERATURE STUDY

Many researchers found that there is lack of consensus on how to define CT, on what kind of skills required to teach CT, and constructing the best procedures and strategies for this teaching. Generally, even now educational projects have not reached common implementations to teach CT. This study reflects on these different views in order to construct a clear understanding of different aspects related to CT.

2.1. Definition of Critical Chinking

An extensive review of literature indicates that some scholars employ CT and high order thinking seemingly. Elliot (2006) is one of the scholars who were interested in studying and using CT and higher order thinking skills in teaching context. However, other scholars have different use of these two concepts and implement them differently according to the proper context and aim Zohar (2004). The connection between "critical thinking," "higher order thinking," "thinking skills" and other terms such as "informal logic," "informal reasoning," "problem solving," "argumentation," "critical reflection," "reflective judgment," and "metacognition" have made them more difficult to apply differently. Scholars also have other issues that led to different understanding like (a) the degree to which CT is related to specific subject, (b) variations between novice and expert thinking in which novices can learn to reflect more like experts, and (d) if CT should be implemented as a process-based practice or a package of skills (Qing, Jing, and Yan, 2010, Keng, 2006, Gambrill, 2005, Thayer-Bacon, 2000). While some scholars have tried to use order on these different terminology (Preiss and Sternberg: 2010), none of them has managed to provide a description, classification, or even a theory could be conventional as ultimate (Martin: 2010, Folsom, 2009, Chenault and Orsello, 2008, Dagli, 2008). The main problem to this disagreement has rested in the grounding of different hypotheses separate fields that is related to this study. These two fields are philosophy and psychology. The current study aims to discuss both views in order to provide a clear picture of the nature

of critical thinking in education. Many philosophers have tried to discuss the quality and the nature of the results or conclusions of CT, such as analyzing justifications and reasoning responses. However, psychologists have focused on cognition procedures and process, the elements and practices implemented to indicate practical learning issues. In addition, psychology has been tested and examined through empirical research process, while philosophy has depended on logical justifications and reasoning to conclude solutions and recommendations. On the other hand, some educators realized the importance of both areas of psychology and philosophy to improve an accurate and reliable teaching methods and procedures of CT (Kuhn: 2005, Giancarlo, 2004, Marzano, Pickering, and Pollock, 2003). The current study reflects on this dilemma and will try to investigate the use of SQ on students' CT.

2.2. Socratic Questioning

The real ancient texts that function as basic foundations of what Socrates really said and did reflected a comprehensive knowledge of how Socrates applied the process of applying "counter-examples" to improve a series of questions that could increase the amount of thoughts, ideas, examples, and views to carry out the dialogue to a maximum advantage. It is clear that discussions and dialogues are the good means to maximize comprehension of the course content. On the other hand, there is no precise written manual that demonstrates how Socratic questioning should be used Paul and Elder (2006). Generally, there is no specific Socratic model where someone can thoroughly try to use in teaching and learning practices. Many scholars believe that the cause behind existence and extension of this approach is the way that interpreted. Sahamed (2004) indicates that Socratic questioning (SQ), having occurred about 2,500 years, has logically developed in its various practices today and the method has better adjusted to suit various purposes. Clearly scholars believe that questioning is an important tool of comprehension in teaching and learning context. As a result, scholars believe that Socratic questioning (SQ) can be adjusted and applied in various ways to different stages of comprehension (Paul & Elder: 2006). As a result, in literature we can underline different acts of applying this method, e.g. the "method of Dialectic "Socratic Method" (Paul & Elder, 2006). It also has been termed Socratic seminar and Socratic dialogue (Brickhouse and Smith, 2007). Keng (2006) considers that over different applications of SQ as unreliable and considers that as misapplication of the pedagogy since only unreal features are applied without the nature of SQ. These misapplications have directed other researchers to consider SQ just open-ended questions and answer procedures (McCoy, 2008).

III. METHODOLOGY

The current study used a quasi- experimental design. Since the current study aims at investigating the use of SQ strategies in EFL reading classes on students CT skills, the researcher believes that conducting an experiment is appropriate for the current study to answer the following questions in this research. 1). To what extent do EFL teachers use different SQ types? 2). Are there any statistically significant differences between mean scores of those who are taught through SQ and those obtained by students who are taught CT skills in a normal setting?2). What is the size effect of SQ on students CT skills? The quantitative data was subjected to certain statistical procedures like frequencies and T-test.

3.1. Classroom Observation

The classroom observation checklist contained twenty-nine items grouped into five categories 1). Inference, 2). Recognition of Assumptions, 3). Deduction, 4). Interpretation, and 5). Evaluation of Arguments (Paull and Elder, 2006). To answer the second question descriptive statistics were used, and the mean of the total number of frequencies for each category was calculated

3.2. Sample of The Study

This study is based on a quasi-experimental design where the researcher investigated the use of SQ strategies on students' CT skills in reading classes. In this design two classes were selected randomly from the public schools in the region and were classified in two groups, one control group and one experimental group. In other words, each class has 30 students. Therefore, 60 students subjected to CT pre-and post-tests to measure the result of the intervention program that related to research question 1 which was based on SQ.All students were general female diploma certificate students (grade 12), aged 17-19. Their first language is Arabic. At the time of the study, they had been learning English as a foreign language for at least 11 years. The researcher also selected a smaller group (10 EFL teachers) to examine research question 2 which reflect teachers' use of SQ inside the classroom by using classroom observation checklist before the intervention stage. Having this in mind, the researcher was able to have a clear context of using CT skills prior to the intervention stag

3.3. Watson-Glaser CT Appraisal Test

This test is intended to collect quantitative data relevant to RQ2 as a pre-test and posttest for the experimental and control groups. This test includes 80 multiple-choice items ranging from 2 to 5 in its five sub-sets. Students were given five scenarios to evaluate the performative function involved in each scenario. The five scenarios offer scores for five different subtests ranging from 0 to 80. The five sub-sets focused on the following: (1) Inference: respondents are supposed to conclude inference among different levels of truth or falsity from given

data. (2) Recognition of Assumptions: respondents are expected to identify unstated assumptions or suppositions implied in given statements or assertions. (3) Deductions: respondents are required to determine if certain conclusions logically follow from the given statements and the data. (4) Interpretation: respondents need to evaluate the evidence and come to a decision if conclusions or overviews are based on the given information. (5) Evaluation of Arguments: respondents are subjected to differentdiscussions that are relevant and well-built, and those that are irrelevant and not properly justified in a particular situation (Watson & Glaser, 2006).

3.4. Intervention

The intervention included 8 reading topics. Each topic was taught and covered in one week (5 hours). SQ strategies used as teaching strategies in each reading topic which cover the five CT skills included in Watson-Glaser CT Appraisal Test. The nature of SQ strategies is based on questions that evoke students' CT, which help the teacher and the students to exchange ideas at higher cognitive level through authentic discussions in the lesson The 8 reading texts along with lesson plans, students' tasks, activities, and related assessment procedures to be used in the intervention phase of 8 weeks were reviewed for its content validity through two major sources. First, these documents were presented to and discussed with five EFL language supervisors and 13 EFL senior teachers in eight workshop sessions that were conducted by the Ministry of Education in Dhofar region. Detailed feedback was collected from the participants through face to face discussion and by recording the discussion in the form of dairy notes. Second, the intervention program documents were sent to three reviewers. The first reviewer was the head of Education Department in Dhofar University, the second reviewer was the chief supervisor of the Supervision Department in the Ministry of Education, and, the third reviewer was an associate professor in Languages Department in Dhofar University. The feedback collected from the two sources was utilized for further improvement.

IV. RESULT AND DISCUSSION

The observation checklist was used to collect data on teachers' actual use of SQ inside classroom. This instrument required a systematic observation process of teachers' use of questions in actual teaching situation. It was adapted from Paul & Elder (2006) checklist. The following is the presentation of the data gathered to answer research question 1.

Table 1: Frequencies of The Use of Different SQ Types as Revealed by The Classroom Observation Checklist (N =10)

No	Domains of CT	Minimum	Maximum	Mean	Std. D.
1	Inference	10	79	35.04	16.57
2	Recognition of Assumptions	7	68	29.25	13.89
3	Deductions	0	36	02.63	07.26
4	Interpretation	0	15	01.13	03.30
5	Evaluation of Arguments	0	8	00.25	01.41

Table (1) shows that the most frequently used question type by the EFL teachers in Oman is Inference with mean value (36); Recognition of Assumptions follow with mean value (29). Deductions are less commonly used questions with mean value (3). Interpretation and Evaluation of Argumentsare found to be the least popular of all question types with average use of only (.25), less than (1) question, respectively. The EFL teachers in this study extensively used PIQ. This result lends support to those of previous studies on the same issue. Shomoossi (2004) found that out of a total of 1628 questions, 1335 have been PIQ (about 82%) and only 293 EQ (about 18%). Also, Long and Sato (1983) found that teachers asked a total of 141 questions, only 24 of which were EQ and 117 of which were PIQ.

The researcher believes that EFL teachers' heavy reliance on the use of PIQ may reflect the realities that classroom context follows traditional teaching roles. In other words, the teacher will be the provider of the information and the student will be the receiver of the information.

Therefore, students will be geared primarily to tests and learn English in an environment where they hardly find opportunities for real interaction and communicative learning.

4.1. Normality Test

Normality test was conducted on the control group and experimental group to maintain normality distribution of the sample. Skewness coefficient for the control group ranged from (0.0) to (0.35), and for the experimental group from (0.0) to (0.58). Both

results are close to (0) which indicate normal distribution of the sample in the two groups. Also, Kurtosis coefficient for the control group ranged from (0.46) to (-1.33) and (-0.39) to (–1.10) for the experimental group which is close to (1) for both the groups. This indicates normal distribution of the sample in the two groups. The researcher also used Kolmogorov-Smirnov test to determine if the two samples related to the control group and experimental group had the similar normal distribution. The Z value was (0.65), and the significance value was (0.78) which means that there was no statistical significance at the level of 0.05.

4.2. Pre-Test Homogeneity and Equivalency

In order to measure the homogeneity and the equivalency of the control group and the experimental group, before implementing the intervention of the variables, the participants' English language achievement score (ELAS), their total grade point average (TGPA), critical thinking skills score (CTSS), and intelligence quotient score (IQS) obtained in the tests conducted prior to the intervention phase were subjected to t-test in order to calculate the differences of the afore-mentioned pre-variables between the two groups. Table (2) shows the differences of the significance between the participants in the two groups.

Table (2) Significance of Differences Between the Two Groups in the Pre-test. N=30

Variables	Experimental Group		Control Group		T test	Sig.
	Mean	Std.Div.	Mean	Std.Div.		
ELAS	79.38	9.95	79.52	9.95	0.05	0.96
TGPA	71.72	68.53	71.95	63.74	0.12	0.90
CTSS	29.28	02.39	29.66	2.92	0.54	0.59
IQS	45.69	06.30	45.17	5.99	0.32	0.75

Table (2) shows that there was no significant difference in the mean scores between the experimental group and the control group in the pre-test of the ELAS, TGPA, CTSS, and IQS. This shows that the two were equal in relation to afore-mentioned test scores prior to the beginning of the experiment.

4.3. Homogeneity and Equivalency of CT skills Test

The researcher conducted t-test in the pre-test stage which reflects the differences between the mean scores of the control group and the experimental group in the five domains, Inference, Recognition of Assumptions, Deduction, Interpretation, and Evaluation of Arguments of Watson-Glasser (2006) CT Test. This was to make sure that the dependent variable which is students' CT skills in both the control group and the experimental group was equal prior to the implementation stage. Table (3) below explains the equality of variances.

Table (3): Equality of Variance Test (students CT skills)

Domains of CT	Control Group (N=30)		Experimental Group (N=30)		t-value	Sig.
	Mean	St. D	Mean	St. D		
Inference	6.0	0.85	5.8	0.85	0.78	0.44
Recognitionof Assumptions	5.8	0.69	6.0	0.76	- 1.27	0.21
Deduction	6.0	0.94	5.6	0.82	1.49	0.14
Interpretation	6.0	0.96	6.0	0.93	0.00	1.00
Evaluation of Arguments	5.9	0.70	5.8	0.76	0.54	0.59
Total	29.7	2.92	29.3	2.39	0.54	0.59

It is evident from the above table that there are no significant differences in the pre-test mean scores of students' CTS for both the control group and the experimental group in the five domains of Watson-Glasser's CT Test. The mean difference of inference domain in both groups was (0.44) which is not significant at (t=0.78). The mean difference of recognition of assumptions domain in both groups was (0.21) which is considered as not significant at (t=-1.27). The mean difference of deduction domain in the two groups was (0.14) which is counted as not significant at (t=1.49). The mean difference of interpretation domain in both groups was (1.00) which is not worthy of considering it as significant at (t=0.00). The mean difference of evaluation of arguments domain in the two groups under examination was (0.59) which is not to be considered as significant at (t=0.54). As a result, both the control group and the experimental group were at equal level in all the five domains of Watson-Glasser's (2006) CT Test. This section is meant to discuss the results relevant to the research question 2, as shown in table (4). The data in this table highlights the differences between mean scores of those who were taught CT

through SQ and those who were taught CT skills in a normal setting. The t-test was conducted to know the differences between the two groups in relation to the five CT domains as specified in Watson-Glasser's (2006) CT test.

Table (4): Significance of Differences Between the Two Groups in the Post-Test-N= 30

Domains of CT	Experimental Group		Control Group		T test	Sig.
	Mean	Std.Div.	Mean	Std.Div.		
Total C.T.	45.69	2.82	29.48	2.34	32.83	0.000
Inference.	9.35	1.23	6.24	1.02	10.43	0.000
Assumptions Recognition	8.66	0.97	5.76	0.87	11.93	0.000
Deduction.	9.21	0.73	5.66	0.86	17.03	0.000
Interpretation	9.03	0.98	6.03	0.73	13.20	0.000
Evaluation	9.45	1.06	5.79	0.86	14.45	0.000

The result of t-test comparisons suggests significant differences between mean scores of those who were taught CT through SQ and those who were taught CT skills in a normal setting. Put precisely, there are significant differences at (0.05) between mean scores of those who are in the first group, who were taught through SQ and those in the other group, who were taught CT skills in a normal setting. This result suggests that teaching CT skills through SQ strategies is beneficial and more productive. Furthermore, t-test results show significant differences between the two groups in relation to the five CT domains. It is evident from the result that those students who were taught through SQ gained higher scores than the ones who were taught CT in a normal setting. The mean scores of the five CT domains show that students who were taught CT using SQ were able to develop and use the target skills in the five domains better than those who were taught in a normal setting. This result clearly demonstrates that the students who were taught CT skills through SQ developed abilities to use CT skills in the five domains at different levels of comprehension. This argument gets desired support from Paul and Elder's (2006) remark that SQ can be adjusted and applied through various techniques at different stages of comprehension. The researcher's argument is consistent with the experimental work of Wenning (2006) in which he proposed a typology, which is similar to SQ that can be considered as a "comprehensive treatment kit" to improve CT skills. Though this result of the study is not directly consistent with Forsyth, Paul, Kelley, and McMillan's (2009) study which reviewed 27 studies that examined the result of different programs and courses on CT skills among college students, and they drew a cumulative conclusion which does not support the view that the application of explicit instructional practices lead to an increase in CT. And, at the same time, they did support the findings that college attendance improves CT skills. The college attendance linked with the improvement in CT reflects the autonomous learning abilities of college students as against the school level students who, in contrast, depend on guided learning and therefore, explicit instructional practice to develop CT skills at school levels remains relevant which finds desired support in Thakur and Al Mahrooqi's (2015, p.126) argument that unlike ordinary thinking which is an inborn human ability, CT needs to be taught through implicit and/or explicit instruction. Thus, as an alternative interpretation of Forsyth, Paul, Kelley, and McMillan (2009) study, the researcher intends to claim that the abilities to use CT skills do develop through explicit instructional practices involving SQ strategies at school levels, where the students have not yet become independent autonomous learners unlike the college students referred to in the afore-mentioned study. The results of Garret's (2006) action research on higher level questioning at SQ clearly demonstrates a significant growth (76%) in the students' ability to construct higher-level thinking in terms of using questions at the upper level of Bloom's taxonomy as a result of explicit instructional input and practice. The most prominent finding relevant to research question 2 that emerged from the statistical analyses of the t-test is that teaching CT skills through SQ strategies is beneficial and more productive. Students who are taught through SQ strategies are better able to develop and use the target skills in the five domains of CT. Another important finding suggests that continuous instructional practice leads to upward and steady growth in the effective use of CT resulting in the internalization and gradual scaffolding of knowledge, skills, and competence of CT. Furthermore, an additional finding strongly suggests that CT skills do develop through explicit instructional practice of SQ strategies at school level, where students have not yet become independent and autonomous learners as against the tertiary level students.

COCLUSION

The findings of the study revealed that EFL teachers in Oman were not able to use CT questions that could enhance students' CT skills during the classroom observation stage. With this in mind, the impact on classroom practice was only one pattern which was teacher domination. Teachers provided knowledge, probed and illustrated the target information. There were no opportunities of real communication or

critical thinking practices. However, after the intervention stage, the most prominent finding relevant to RQ2 that emerged from the statistical analyses of the t-test is that teaching CT skills through SQ strategies is beneficial and more productive. Students who are taught through SQ strategies are better able to develop and use the target skills in the five domains of CT. Another important finding suggests that continuous instructional practice leads to upward and steady growth in the effective use of CT skills resulting in the internalization and gradual scaffolding of knowledge, skills, and competence of CT. Furthermore, an additional finding strongly suggests that CT skills do develop through explicit instructional practice of SQ strategies at school level, where students have not yet become independent and autonomous learners as against the tertiary level students.

REFERENCES

[1] Brickhouse, T. & Smith, D. (2007). Socrates on How Wrongdoing Damages the Soul. Available on legacy. Lclark.edu/ ndsmith/45lsyl.htm.

[2] Chenault, T. &Orsel, E. (2008). An Act of Translation: The Need to Understand Students' Understanding of Critical Thinking in the Undergraduate Classroom. The Journal of Effective Teaching, 8 (2), 5-20.

[3] Elliot, A. (2006). Hierarchical Model of Approach-Avoidance Motivation. Business Media, Inc.30 (2), 111-116.

[4] Folsom, C. (2009). Teaching for Intellectual and Emotional Learning (TIEL): Bringing Thinking and Moral-Ethical Learning into Classrooms, Springer Publishing Company, vol. 5, no. 1, pp. 1-16

[5] Forsyth, D. R., Story, P. A., Kelley, K. N. & McMillan, J. H. (2009). What causes failure and success? Students' perceptions of their academic outcomes. Social Psychology of Education, 12(2), 157-174.

[6] Gambrill, E. (2005). Critical Thinking in Critical Practice: Improving the Quality of Judgment and Decisions. New Jersy: Wiley & Sons Press.

[7] Garret, W. (2006). Developing Sixth Grade Gadflies: An Action research project presented to the Shawnee Mission Board of Education. Bluejaket-Flint Elementary.

[8] Gian Carlo, C. (2004). Assessing Secondary Students' Dispositions towards Critical Thinking: Development ofthe California Measure of Mental Motivation, 64 (2). 347-364.

[9] Kelly, K. (2009). Assessing Students' Critical Thinking Performance: Urging for Measurement Using Multi-Response Format. Journal of Science direct, 4 (1), 70-76.

[10] Keng, L. (2006). Critical Thinking and Socratic Inquiry in the Classroom. Paper presented at the Joint ERA/AARE Conference, Singapore 1995. Australian Association for Research in Education.

[11] Kuhn, D. (2005). Education for Thinking, Library of Congress Cataloguing-in-Publishing Data, Harvard College language learners. The Modern Language Journal, 83, 193-201.

[12] Martin, R. (2010). Bells that Still Can Ring: Systems Thinking in Practice. In: 1st International Workshop on Complexity and Real World Application Using the Tools and Concepts from Complexity Science to Support Real World Decision-Making Activities, 21-23, July, England, UK.

[13] Marzano, R., Pickering, D. & Pollock, J. (2003). Classroom Instruction that Works: Research-Based Strategies for Increasing Students' Achievement, Library of Congress Cataloging-in-Publication Data, USA.

[14] Ministry of Education, Oman (2008). Basic Education, English Language Curriculum Framework. Curriculum Development.

[15] Paul, R., & Elder, L. (2006). The Art of Socratic Questioning: The Foundation for Critical Thinking, www.criticalthinking.org. retrieved on 20/05/2016.

[16] Preiss, D. & Sternberg, A. (2010). Innovation in Teaching Psychology: Perspectives on Learning, Teaching, and Human Development, Springer Publishing Company, pp. 357-395.

[17] Qing, Z., Jing, G. & Yan, W. (2010). Promoting Pre-service Teachers' Critical Thinking Skills by Inquiry-Based Chemical Experiment. Innovation and Creativity in Education, 2 (2), 4597-4603.Quarterly, 29, 144-155.

[18] Sahamid, H. (2004). Socratic Questioning in The Teaching of Short Stories: An Action Research Study. The Reading Matrix, 4 (2), 1-12.

[19] Shomoossi, N. (2004). The Effect of Teachers' QuestioningBehavior on EFL Classroom Interaction: A Classroom Research Study.The Reading Matrix. Vol. 4, No. 2, 1-9.

[20] Thayer-Bacon, B. (2000). Transforming Critical Thinking: Thinking Constructively: To Enhance Thinking Skills, In: The Teaching of Psychology, Teachers College Press, Columbia University Taylor & Francis e-Library, pp.91-120.

[21] Watson, G. &Glasser, E. (2006). Watson-Glasser critical thinking appraisal short form manual. San Antonio, TX: Harcourt.

[22] Wenning, J. (2006). Engaging Students in Conducting Socratic Dialogues: Suggestions for Science Teachers. Available on wenning@phy.ilstu.edu.

[23] Zohar, A. (2004). Higher Order Thinking in Science Classroom, Students' Learning and Teachers' Professional Development. Kluwer Academic Publishers.

★★★