

AMERICAN UNIVERSITY OF ARMENIA

College of Humanities and Social Sciences

Exploring an Impact of Teaching Argumentation on the Development of Writing Skills

A thesis submitted in
partial fulfillment of the requirements for the degree
Master of Arts in Teaching English as a Foreign Language

By

Armine Petrosyan

Rubina Gasparian, Ph.D., Adviser

....., Reader

Yerevan, Armenia

May 13, 2018

We hereby approve that this thesis

By

Armine Petrosyan

Entitled

Exploring an Impact of Teaching Argumentation on the Development of Writing Skills

Be accepted in partial fulfillment for the requirements of the degree

Master of Arts in Teaching English as a Foreign Language

Committee on the MA Thesis

.....
Rubina Gasparyan, Adviser

....., Reader

.....
Irshat Madyarov

MA TEFL Program Chair

Yerevan, Armenia

May 13, 2018

TABLE OF CONTENTS

List of Tables	v
List of Figures	vi
Abstract	vii
Chapter One: Introduction	1
1.1 Statement of the Problem	1
1.2 Purpose of the Study	2
1.3 Research Questions	2
1.4 Overview of the Study.....	2
Chapter Two: Literature Review	3
Chapter Three: Methodology.....	11
3.1 Research Design	11
3.2 Participants	11
3.3 Instruments	12
3.4 Procedures	12
3.5 Data Collection and Processing	15
Chapter Four: Data Analysis and Results	17
4.1 Quantitative Data Analysis	17
4.2 Qualitative Data Analysis	21
Chapter Five: Discussion and Conclusion	23
5.1 Quantitative Impact of the Treatment: Factors that Did Not Work	23
5.2 Qualitative Impact of the Treatment: Factors that Worked	26
5.3 Limitations	32
5.4 Delimitations	33
5.5 Recommendations: What to do Differently	34
References	36

Appendices

Appendix A. Scoring guide	40
Appendix B. Treatment syllabus	41
Appendix C. Scores on pre- and post-treatment essays	49
Appendix D. Parametric tests results	50
Appendix E. Non-parametric tests results	54
Appendix F. Cronbach's alpha test results	57
Appendix G. Observation-reflection journal	58

LIST OF TABLES

Table 1. <i>Inter-rater reliability: Correlation between the scores of all essays of the two groups</i>	17
Table 2. <i>Scores on pre- and post-treatment essays of comparison and treatment groups</i>	19
Table 3. <i>The “Three Ms” of the essay scores</i>	19
Table 4. <i>Lengths of the essays and the use of liking words</i>	27

LIST OF FIGURES

<i>Figure 1.</i> Comparisons of between the groups	16
<i>Figure 2.</i> Consistency between the scores of the two raters	18

Abstract

Argumentation skills are among of the most acclaimed 21st-century skills. When combined with language skills, arguments become a powerful tool. Although there is a consensus in the literature about the benefits of using arguments and debates in foreign language classrooms, little, if any, research exists in Armenia that would provide empirical data in support of possible connections between teaching argumentation skills and English language competency. The purpose of this mixed-methods study is to explore possible connections between explicit instruction in argumentation and writing skills of Armenian EFL students. The quasi-experimental design was originally used. Two groups – comparison and treatment groups– of low-intermediate-level students of a language school constitute the sample. The scores on *pre-* and *post-*experiment in-class composition writing provided the quantitative data for analysis. The results of the quantitative analysis revealed no statistically significant differences between the two groups after the treatment period. However, the qualitative analysis that followed indicated some tentative signs that gave hope for further efforts in search of the above-mentioned connections.

Keywords: explicit instruction, argumentation skills, language learning, language skills, writing skills, debate, reasoning

CHAPTER ONE: INTRODUCTION

In the 21st century, which is all about power and who-gets-what politics, the possession of yet another powerful tool - argumentation skills - can be a decisive factor in many circumstances. Instruction in argumentation skills and debate is widely used in science education and for instruction in the students' first language in many Western countries (Lin & Mintzes, 2010). However, using argumentation and the structured academic debate format in EFL classes is an innovative approach to foreign language instruction (Brown, 2009; Morse, 2011).

The scholarly sources reviewed in the literature clearly favor the idea of using argumentation and debate in foreign language instruction; but there seem to be few research projects, let alone empirical data, that shed light on the possible relationships between the instruction in argumentation skills and the achievements of EFL students. Still less, if any, such research and data exist in Armenia, which lends a degree of significance to this study. However, the significance is of limited nature: the results will only have meaningful interpretations when applied to the participants and, at best, to the language school they come from. Having said that and bearing in mind that this research is, by all evidence so far, the first such attempt in Armenia, other similar schools in the country may benefit from familiarizing themselves with the methodological approach and the discussion of the results.

This study has also important pedagogical implications. How exactly the EFL students of the treatment group was instructed in argumentation skills is a factor that was closely reflected on during and after the actual treatment.

1.1 Statement of the Problem

Based on the evidence examined for this study, there is an acknowledged scarcity of research worldwide on a relationship between instruction in argumentation and English language writing skills. Still less, if any, such research exists in Armenia.

1.2 Purpose of the Study

The purpose of this paper is to explore whether there are relationships between the EFL students' exposure to explicit instruction in argumentation and their progress in gaining English writing skills.

1.3 Research Questions

RQ1: To what extent does explicit instruction in argumentation impact the writing skills of the Armenian English language learners?

RQ2: What evidence of the impact of explicit instruction in argumentation on the writing skills of the Armenian English language learners does qualitative data analysis reveal?

In the above research questions, explicit instruction in argumentation means teaching the two argumentation skills, namely, guarding and discounting the premises and the basic types of argument markers, namely, reason, conclusion and contrast markers.

1.4 Overview of the Study

The remainder of this paper consists of the following parts. It starts with a literature review that focuses on the major approaches to teaching of argumentation skills (in general, and in L1 and L2 classes), on pedagogical implications and challenges, such as the learners' age, the culture (or the absence thereof) of academic debates, the use of controversial topics, and the relevance of these issues to the present study. The literature review is followed by a chapter on methodology that draws on the relevant studies found in the literature.

Particularly, the methodology chapter discusses how (based on what existing studies) the treatment syllabus was developed, the treatment group was chosen, what was the role of the teacher, and the types of the collected data. Next comes the description of the process of quantitative and qualitative data collection and the results that the data analyses revealed.

Then the discussion of the results focuses on the impact that the treatment syllabus had on the

students' writing skills: factors that did and did not work during the experiment. Limitations that follow primarily highlight the challenges faced in the course of this study and particularly the reasons of why some factors did not work during the implementation of the teaching syllabus. A few unassuming recommendations conclude the paper.

CHAPTER TWO: LITERATURE REVIEW

This chapter focuses on major approaches to teaching argumentation – in general, in L1 and L2 classes, and particularly in L2 writing classes. It then goes on to discuss the pedagogical aspects of teaching argumentation. Throughout the discussion, it also shows how the studies and approaches found in the literature are relevant to the present research.

The scholarly literature approaches the concept of argumentation skills from two global perspectives. On the one hand, the researchers view teaching and learning argumentation as an end in itself that can be achieved directly by explicit instruction in argumentation (Rapanta et al., 2013; Reznitskaya et al., 2007; Rybold, 2006) and indirectly through, for example, using the electronic mail exchange for practicing argumentation (Marttunen, 1997), through eliciting questions from the students in science classrooms (Chin & Osborne, 2010), or through instruction in socio-scientific issues (Dawson & Venville, 2010; Lin & Mintzes, 2010). As an end in itself, learning argumentation skills depends on various factors, such as the activation of an “argument schema” that pre-exists in individuals' minds (Reznitskaya et al., 2007) or the learners' “ability” as measured by an IQ test (Lin & Mintzes, 2010). On the other hand, argumentation skills are viewed as a tool for pursuing something else – from promoting learning in general (Horn, 2008) to developing more specific skills, such as critical thinking (Morse, 2011), learning language skills, grammar, and vocabulary (Moss, 2011), writing skills (Johns, 1993), or developing communication skills (Morse, 2011; Zhang & Wang, 2012).

This paper focuses on the second global approach to teaching argumentation, which views it as a means of accomplishing a goal or goals. Among various goals that can be achieved through teaching and learning argumentation skills, this study singles out language learning, particularly, writing skills. However, the first global approach, i.e., teaching argumentation for its own sake, is also relevant to the present study because of its important pedagogical implications that were instrumental in developing the treatment syllabus. In other words, this research adopts a dual approach to teaching argumentation skill. The specific pedagogical implications and the relevant literature will be discussed later in this section. Also, not to digress from the literature review, it is worth mentioning here only briefly that the reasons behind this study's dual take on teaching argumentation skills will be discussed in the methodology section.

Following the dual approach to teaching argumentation adopted in this study, the literature review will now look for the two groups of sources on teaching argumentation in greater detail – sources that are about teaching argumentation for linguistic purposes in L1 and in L2 classes, and those that are about pedagogical implications of teaching argumentation *per se*, for its own sake.

While the literature that looks into various aspects of the link between argumentation and language skills is abundant in general, it shows one important characteristic: it seems much more prolific in exploring the influence of argumentation skills on students' first language than when it comes to exploring its influence on students' second language. In Western countries, many universities include argumentation and debate in the courses they teach, e.g., in geographical education in the United Kingdom (Morgan, 2006). This pursues "the dual goals", i.e., teaching reasoning/argumentation skills and "improving students' command of their first language" (Brown, 2009, p. 536). Sinnott-Armstrong & Neta (2011), too, emphasize a strong association between argumentation and language skills by stating that

“the material of argument is language” (Lecture 4), i.e., arguments are made of language. Likewise, for Song et al. (2013) the link between argumentation skills and language is so strong that they use the former as a learning progression for teaching and evaluating English language arts. Painter et al. (2003) also were apparently driven by their interest in the same link when they designed debating activities for their distance learning course “to bring to students’ attention linguistic features of argumentation” (p. 162). Still another study is by Hickey & Watson (2014), who emphasized teaching argumentation to English learners as their L1. The authors discussed the strong relationship between language and argumentation in terms of The Common Core State Standards that “forefront [...] argument as a writing genre” (p. 122). It bears repeating that in all these studies, the English language was the students’ L1.

From the above-cited sources, it can be seen that the use of argumentation and debate for improving language skills in L1 classes is common practice. However, it does not seem to be as common in L2 classes. Moreover, such practices are viewed as “quite innovative” (Morse, 2011, p. 111). Brown (2009), too, mentions that debates are seldom, if ever, to be found in foreign or second language syllabi. These opinions can at least partly explain the relative scarcity of research and, more importantly, the lack of empirical data about the nature of the relationship between argumentation skills and proficiency in L2, particularly in L2 writing. One such rare piece of evidence is the research conducted by Brown (2009). Because among the studies in the literature investigated so far this particular research comes closest to what is central for the present paper, it deserves a detailed description here.

The author studied a group of 14 native English speaking university students enrolled in a Russian-language class called “Global Diplomacy and Debate.” The sample distinguished between six proficiency levels: intermediate-mid and -high, advanced -low, -mid and -high, and superior. These proficiency levels were determined by using the ACTFL

(American Council on the Teaching of Foreign Languages) scale. The students completed pre- and post-oral proficiency interviews and written proficiency tests. The pre- and post-test ratings were compared. The results revealed a statistically significant difference, but only for the “advanced-high” group. The author explains this by the theory of Vygotsky’s (1978) Zone of Proximal Development (ZPD) by arguing that “the difficulty of assignments in oral debates complemented the ZPD of participants who began at the advanced-high level more than the ZPD of participants who began at the advanced-mid level” (pp. 545-546). This explanation, however, is not based on the statistical data collected. The data in Brown’s (2009) study was about the students’ proficiency levels and the ratings of their performance; no empirical data was collected on the ZPDs of the students. Therefore, the authors’ explanation of the “advanced-high” group’s better performance in terms of ZPD, although possible in theory, is rather speculative. Besides, the design of Brown’s research seems to be complicated (e.g., division of the small sample of only 14 students into six proficiency subgroups, including the three sublevels within the “advanced” level, followed by yet another “superior” levels of proficiency). And yet, being one of the rare research projects that directly studies the connection between argumentation skills and L2 written proficiency (which constitutes the central idea of the present paper), Brown’s (2009) study is particularly valuable because it sheds light on the important issues in this seemingly underexplored area of inquiry. Moreover, the author’s statement that “innovative curricular design can equal if not exceed uptake that occurs in extended immersion environments” (p. 534) shows his researcher’s enthusiasm about the effectiveness of using argumentation and structured debate in foreign language instruction. To come across this attitude among the rare research on the exact topic of one’s own professional interest gives a feeling one would have if another civilization was found in the Universe.

The literature cited so far in this review suggests that teaching argumentation and debate for improving student's language skills is by far more common in L1 than in L2 classes. But the present literature review will not stop here. It will go on to look into the pedagogical implications of teaching argumentation *per se*, which are central to this research: How exactly argumentation skills are taught, especially to teenagers who are the target age group of the study. The literature on this important issue suggests several approaches.

One approach is to teach argumentation skills through instruction in socio-scientific issues - the SSI instruction (Kim & Lee, 2017; Lin & Mintzes, 2010). According to the authors, facing the familiar but contentious issues in the classroom causes the students (high-school in the first study and 6th-graders in the second) to stretch their learning abilities to be able to participate in the argumentation-related tasks set around these issues. The present study adhered to this very principle by using familiar and contentious issues in classroom debates for teaching argumentation (Appendix B. Treatment Syllabus).

Another author (Dundes, 2001) approached the issue of teaching argumentation from the perspective of the classroom size. He suggested that groups be small so that students could be maximally involved in class activities. For this, he divided his relatively big class of 30 students into smaller groups to ensure all group members' participation in the discussion. In the present research, too, the treatment group was small enough to have three debaters on each side of the discussion and the rest acting as judges; the maximum involvement was thus guaranteed. Dundes (2001) also used two criteria for evaluating the debates in his small groups – performance of the debaters and the students' feedback on that performance. Again, this same approach was used during the classroom debates in the present study.

Other authors (e.g., Sinnott-Armstrong & Neta, 2011) advocate a substance-based approach to teaching argumentation that was also used in the present study. The authors singled out a few basic elements (substances) of argumentation, such as argument markers,

including introduction, reason, conclusion and contrast markers, and a few basic moves in argumentation, such as guarding and discounting the premises. According to the authors, the debater guards the premise of his argument when he deliberately makes it weaker and therefore harder for his opponents to disagree with it. By the same token, the debater discounts the premise of his argument when he openly cites the possible objection by his opponent in order to anticipate or even to prevent it. Around these basic elements of argumentation, the authors developed a full-fledged course of teaching argumentation. The treatment syllabus of the present study, albeit much shorter, was also developed around these basic elements of argumentation (Appendix B).

The above discussed pedagogical implication of teaching argumentation informed the development of the treatment syllabus, but we did not lose sight of the research questions of this study that reflect the main idea behind teaching argumentation skill, namely, its impact on the students' English language skills. One full-fledged theory of second language acquisition (SLA) widely discussed in scholarly sources will take this literature review further. It is called Processability Theory (PT) (Pienemann, 1998, 2015). PT claims (and empirically supports the fact) that teaching will not result in actual learning unless the learners' interlanguage is ready to process what is being taught. In other words, what is being taught should be teachable as well as learnable. And this adds yet another subtlety to the discussion of pedagogical implications of teaching argumentation. Remembering here that arguments are made of language (Sinnott-Armstrong & Neta, 2011) and trusting the Processability Theory which is generally acknowledged even by its opponents (e.g., De Bot et al., 2007), it is easy to deduce one more pedagogical implication of teaching argumentation, namely, that in L2 classes it should be taught in a language that is learnable and teachable, i.e., processible, for learners. As will be shown later in this paper (Section 5.2.2), there was one particular episode observed by the researcher in one of the classroom

debates that seemed to fully agree with this fascinating claim of the Processability Theory and with this particular pedagogical implication of teaching argumentation in an L2 classroom.

In the light of PT's concept of teachability and continuing along the lines of pedagogical implications, it is interesting to mention two more studies relevant to the present paper. One study (Conner et al., 2014) addresses the teacher's role in teaching argumentation. The author focuses on "collective argumentation as an important part of classroom discourse where teacher is the moderator" (p. 404). It is when this "collective argumentation" is set in a foreign language classroom with the teacher as moderator, as was the case in the present study, that the above-mentioned teachability component of Pienemann's processability theory is relevant for consideration. For in such a setting, the teacher's role goes beyond the technical moderation of the debate; the teacher should also make the language of argumentation understandable for the students, for example, by bringing examples of argument markers (Sinnott-Armstrong & Neta, 2011) or choosing the debate topics from the students' regular textbook, as was done during the classroom debate session in our treatment group. The second study (Horn, 2008) concerns the managing of disagreements in the classrooms where the students are children who "bring in their own personal understanding of how disagreements are managed" (p. 100). The study introduces one specific structure of the classroom discussion, called *accountable argumentation*, as a form of participation that promotes learning through argumentation. The author uses the term "accountable argumentation" because "it behooves participants to be responsible for many elements of the discussion" (p. 104). She then defines certain "norms and expectations" of participation, such as "value disagreements and understand that they may not be resolved", "have a [...] position in a discussion", "respectfully respond to others' positions" (p. 104). Rules very similar to these norms were used during the classroom debates as part of the treatment syllabus in the

present study. For example, the classroom debate rubric featured “Respect for the other team” as a judgment criterion to determine the winner (re: Appendix G: Observation-reflection journal, entry of January 16, 2018). Furthermore, one of Horn’s (2008) norms, the norm “value disagreement”, was the teacher’s responsibility in the implementation of the treatment component of this research; during each debate session, the teacher emphasized that disagreements are normal and indispensable in debates and need not even be solved to determine the outcome of the debate. Also, the students were assigned roles during the debates, similar to Horn’s norm “have a position in a discussion” mentioned above. One essential difference between Horn’s (2008) study and the present research is that in Horn’s classroom the disagreements were managed/moderated by the students themselves, whereas in the present research it was mostly the teacher’s responsibility. This issue is addressed in more detail later in this paper (Section 5.3. Limitations). Thus, although not immediately language-related, these two studies were helpful for the present research in understanding how to deal with contentious issues discussed in the classroom during the structured academic debates.

Summarizing the literature review, it is important to reiterate that teaching argumentation skills and debates is widely used in many countries for instruction in L1 and are rarely used for instruction in L2. The researchers view the latter approach as innovative. The research projects in this innovative direction that investigate the connection between argumentation skill and L2 proficiency are few and far between. By all evidence, the probability is high that such projects have not yet been done in Armenia.

CHAPTER THREE: METHODOLOGY

This chapter presents the major components of the methodology used in the study: descriptions of the research design, of the two groups of participants of the study, the instruments used, the main features of the treatment, the method of assessment of the essays, and the characteristics of the collected data.

3.1 Research Design

This study uses the mixed-method approach and has both quantitative and qualitative components. It was originally designed as quantitative quasi-experimental. There were two groups of learners – comparison and treatment. Both groups were pre- and post-tested with in-class essay writing. In the course of eight weeks, twice a week, both groups followed their regular syllabus. During the same period, one of the two groups, the treatment group, was additionally treated to a series of argumentation-related activities, such as in-class oral debates/discussions and explicit instructions in argumentation skills. The pre- and post-test essay scores for the two groups were compared to see whether there were statistically significant differences between the groups' performances before and after the treatment period.

After the quantitative analysis had revealed no statistically significant differences between the two groups, the qualitative component was added to the research design to dig deeper in search of some tentative impact of the treatment. The three sources of qualitative data were the students' essays, the researcher's observation of the classes, and the students' feedback in a focus group discussion.

3.2 Participants

The two groups of participants were selected from among EFL speaking students of a language school. The age range of the students is 12-16 years. All students were native speakers of Armenian; their knowledge of the English language was low-intermediate level.

The number of participants, or the sample size was: 16 students in the comparison group and 11 students in the treatment group.

3.3 Instruments

The students of both groups wrote *pre-* and *post-*test in-class essays that were graded using a holistic rubric (Appendix A). The grades were compared to capture possible statistically significant differences between the groups after the treatment.

The students' writing skills were measured using a direct method of assessment, i.e., by rating their pre- and post-treatment essays, because "direct measures have greater construct validity" (Weir, 1993, p. 133). The inter-rater reliability, or the consistency of the essay scores, was ensured by having the essays rated twice, i.e., using the rates of two independent raters (Hughes, 1989). The final scores of the essays were the simple arithmetic averages of the two rates.

3.4 Procedures

3.4.1 Selection of the Treatment Group

The group size was the decisive factor in selection of the treatment group to have all the students maximally involved in class activities, which echoed the studies (Dundes, 2001) discussed in the literature review. The treatment group was determined before any scoring was done on the pre-tests of the groups. The pre-test grades did not affect the decision.

3.4.2 Grading Procedure

The students' pre- and post-treatment essays were graded twice by two independent graders. The grades were then averaged and used as the final scores for the essays. This procedure ensured inter-grader reliability of the scores.

3.4.3 Teacher

The teacher was the same for the two groups - to make the conditions as similar as possible. The only difference between the groups in relation to the teacher and the syllabus was the instruction in argumentation that is reflected in the treatment syllabus.

3.4.4 Treatment Syllabus

Both the comparison and the treatment groups were taught using their regular syllabus and the textbooks which were identical for the two groups. The students in the treatment group were additionally taught using the treatment syllabus (Appendix B). The treatment syllabus consisted of two components. One component included the basic techniques of argumentation and was developed based on the research discussed in the literature review chapter of the present paper (Sinnott-Armstrong & Neta, 2011). The other component was the actual conduction of structured debates which also drew on the existing research discussed earlier (Dundes, 2001). The entire treatment was spread around, and conducted amid, the students' regular syllabus. The treatment syllabus was divided into two parts: ten hours for conduction of in-class structured discussions/debates and seven hours for teaching the basic techniques of argumentation through explicit instruction in argumentation skills: (a) identifying argument markers; (b) guarding and discounting the premises of an argument; and (c) using introduction, reason, contrast and conclusion markers in discussions (Sinnott-Armstrong & Neta, 2011). These techniques were taught and recycled by applying them to different contentious topics taken from the groups' regular textbook, suggested by the students themselves or, on rare occasions, by the teacher.

3.4.5 Teaching Argumentation Skills

As was mentioned in the literature review section, this research takes a dual approach to teaching argumentation skills, i.e., as a means of achieving the goal of improving writing skills and as an end in itself. Because the present research looks for possible relationships

between teaching argumentation and improving the students' language skills, the purpose of teaching argumentation is viewed here as a means of achieving that goal. At the same time, it is also an end in itself because, in order to answer the research questions, the teaching of argumentation was explicit, i.e., it was taught "as though" for its own sake, i.e., without any attempt to connect it with writing skills. The purpose of this approach was to avoid the Hawthorn effect, which is a type of motivational effects of observed experiments (Dornyei, 2007). The students of the treatment group were not supposed to know about the experiment, i.e., that they were taught the elements of argumentation for improving their writing skills. If the students were made aware of the similarities between, for example, organizing a constructive speech in a debate, and organizing the structure of an essay, it would amount to teaching them writing skills too, which would compromise the results of the experiment. In other words, telling the students the real purpose of teaching them the argumentation skills would make the impact of the treatment on writing skills look much stronger. However, the first research question then could have looked into a relationship between teaching writing skills and improving writing skills, which would be tautology. For these reasons, to the students, and in some sense even in reality, argumentation was taught "as though" for its own sake. This was also a big challenge that turned out to be a limitation (re: Section 5.3). But even at the risk of facing more limitations this condition was strictly observed during the treatment.

3.4.6 Researcher

The researcher observed most of the classes of the comparison group and all classes of the treatment group and kept an observation-reflection journal (Appendix G). She also sent regular emails to the students of the treatment group before each class. These emails routinely contained three major points: useful links for the upcoming debate, a reiteration of the structure of the constructive speeches (both negative and affirmative), i.e., "tell what you're

going to tell, then tell it, and then tell what you've just told", and a friendly reminder of the importance of homework research, at least a quick googling of the topic on the Internet.

3.5 Data Collection and Processing

The quantitative data consisted of the scores that the students of both groups received for the essays they wrote at the beginning and at the end of the term. The pre- and post-test scores were compared to see whether there were statistically significant differences between the groups.

The collected data were processed and analyzed with the use of the Statistical Package for Social Sciences (SPSS). Because of the small sample size, the non-parametric statistics were applied – the Mann-Whitney U-test and the Wilcoxon Single Rank Test – which are the non-parametric analogues of the Independent-sample and Pared-sample t-tests, respectively (Dornyei, 2007, p. 230). These tests were used in search of significant differences between the pre- and post-test scores for each group, as well as between the post-test scorings of the two groups. Figure 1 below visualizes these “vertical” and “horizontal” comparisons.

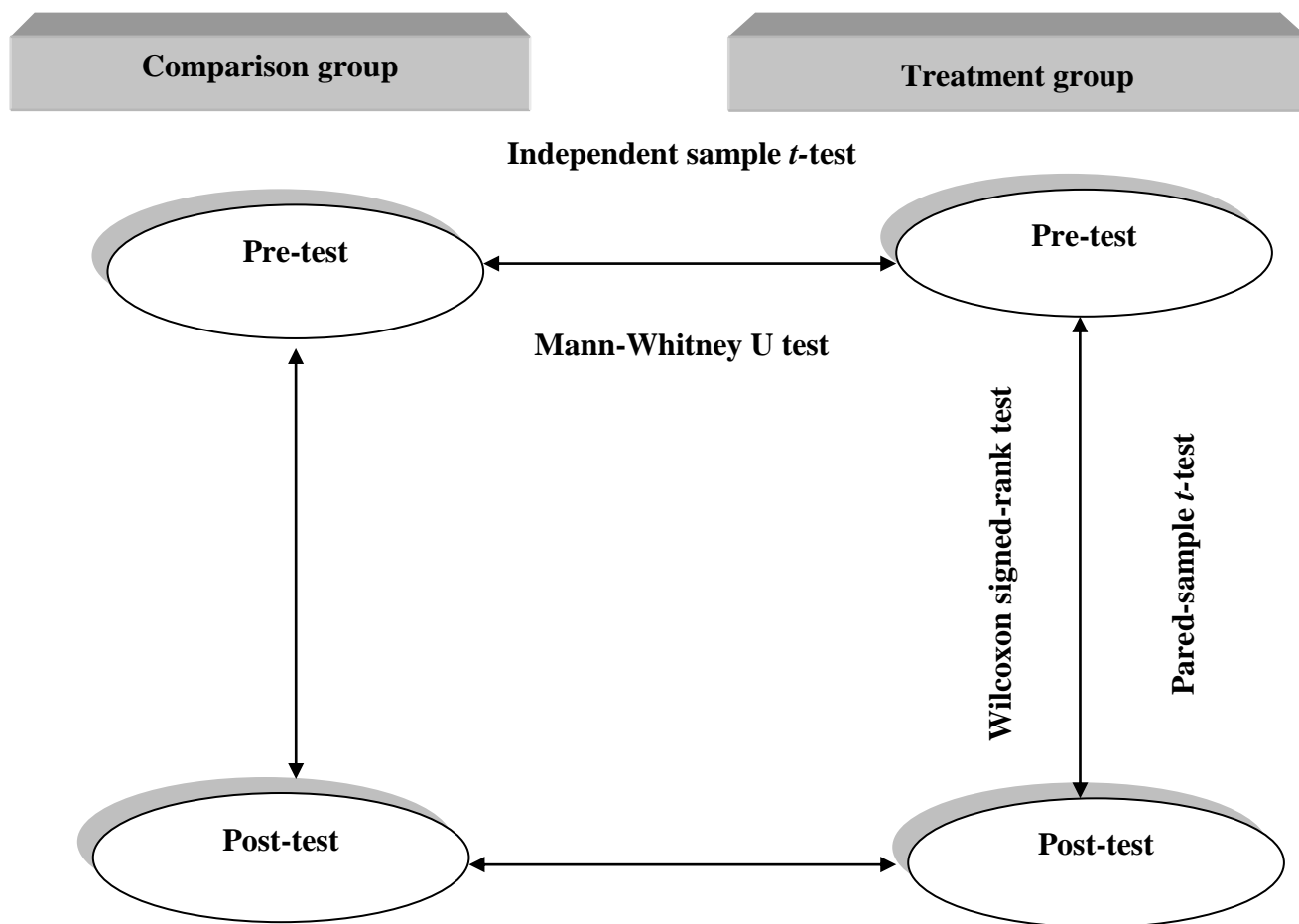


Figure 1. Comparisons of between the groups

CHAPTER FOUR: DATA ANALYSIS AND RESULTS

This chapter presents the process of both quantitative and qualitative data analyses. The quantitative data was the major source of analysis in search for the answer to the first research question. In addition, qualitative data was also collected in search of some features of the treatment that the quantitative data leveled out and also to answer the second research question.

4.1 Quantitative Data Analysis

As was mentioned in chapter 3 on methodology, the quantitative data consisted of the grades of the essays the students of the two groups wrote before and after the treatment. To ensure inter-rater reliability, the essays were graded twice, by two independent graders. Table 1 shows that there was a statistically significant positive correlation between the scores given by the two graders on the pre-test essays of both groups ($r = .48, p < .001$). In other words, the grades of the two graders were consistent.

Table 1. *Inter-rater reliability: Correlation between the scores of all essays of the two groups*

		All_essays_rater_1	All_essays_rater_2
All_essays_rater_1	Pearson Correlation	1	.477**
	Sig. (2-tailed)		.000
	N	52	52
All_essays_rater_2	Pearson Correlation	.477**	1
	Sig. (2-tailed)	.000	
	N	52	52

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

Figure 2 displays the high level of consistency between the scores of the two graders given on the essays of both the comparison and the treatment group.

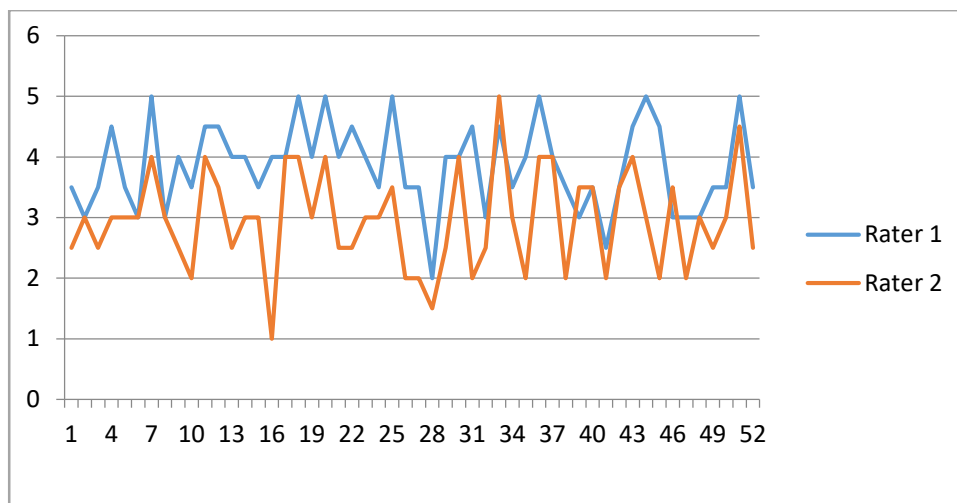


Figure 2. Consistency between the scores of the two raters

To check the inter-rater reliability further, the Cronbach's alpha test was also run (Appendix F). The test gives "a measure of internal consistency" (Larsen-Hall, 2010, p. 391). It returned the value of Cronbach's alpha coefficient $\alpha = 0.7$, which is the lowest acceptable level of internal consistency for the scores given by different raters (Larsen-Hall, 2010, p. 171). Together, Table 1, Figure 2 and the Cronbach's alpha coefficient show a high level of inter-rater reliability in grading of the essays.

Inter-rater reliability being thus ensured, the scores of the two raters on each essay were averaged, and all the subsequent statistical analysis was based on these average scores (Appendix C). Table 2 below shows the descriptive statistics of the four sets of scores. It also shows two important facts: i) that the mean post-treatment scores for both the comparison and the treatment groups dropped, and ii) that the four sets of scores are similar in terms of their descriptive statistics.

Table 2. *Scores on pre- and post-treatment essays of comparison and treatment groups*

Comparison group						Treatment group					
Pre-treatment			Post-treatment			Pre-treatment			Post-treatment		
<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
3.4	0.52	15	3.3	0.78	15	3.6	0.65	11	3.5	0.67	11

Additionally, Table 3 shows that the values of the three major descriptive parameters – the “Three Ms” for the four sets of scores – are close to each other. The relatively small standard deviations indicate that the distribution of values in the four sets of scores are close to their mean values, i.e. the values are not widely spread around their means.

Table 3. *The “Three Ms” of the essay scores*

n/n	Comparison group		Treatment group	
	Pre-treatment	Post-treatment	Pre-treatment	Post-treatment
Mean	3.4	3.3	3.6	3.4
Mode	3.0	3.3	3.5	3.3
Median	3.3	3.3	3.5	3.3

These characteristics of the four variables indicate that their distributions are near-normal, which made it possible to use t-test statistics to compare the groups. Two Independent-sample t-tests and two paired-sample t-tests were used in search for statistically significant differences between the means of these four variables. Appendix D contains the SPSS output tables for these four t-tests.

- Independent-sample t-test 1: to compare the means of the pre-test scores of the comparison and the treatment groups
- Independent-sample t-test 2: to compare the means of the post-test scores of the comparison and the treatment groups
- Paired-sample t-test 1: to compare the means of the pre- and post-test scores of the comparison group

- Paired-sample *t*-test 2: to compare the means of the pre- and post-test scores of the treatment group

The four *t*-tests mentioned above revealed statistically non-significant differences between the mean values of the score distributions. The first comparison showed that there was no significant difference between the two groups of participants before the treatment ($t = 0.86$, $df = 24$, $p > .05$), i.e., that the groups were identical at the beginning of the experiment, in terms of their essay scores. This was a much looked-for result because the more similar the two groups are at the beginning of an experimental study the better (Dornyei, 2007). The second comparison, however, showed that the groups were also similar after the treatment ($t = 0.52$, $df = 24$, $p > .05$). This latter result already was a signal that toward the end of the experiment the treatment group would not be much ahead of the comparison group. Indeed, the fourth *t*-test confirmed that the treatment group did not show statistically significant improvement as a result of the treatment ($t = 0.78$, $df = 10$, $p > .05$). The SPSS outputs of the *t*-tests are given in Appendix D.

The results of the *t*-tests were unequivocal in that they revealed no statistically significant differences. However, because the two groups of participants were small, for all four comparisons the non-parametric statistical tools were also applied, in yet another attempt to find statistically significant differences between the groups, since “in some cases the new coefficients ... [can be] actually higher” (Dornyei, 2007, p. 228). The non-parametric tests are the analogues of the parametric *t*-tests normally used for larger groups. The first two comparisons used the Mann-Whitney U test, the non-parametric analogue of the Independent-sample *t*-test; for the second pair of comparisons, the Wilcoxon Signed-Rank test was used, the analogue of the parametric paired-sample *t*-test. The comparisons between the groups made with the help of non-parametric test, just like their parametric counterparts, did not reveal statistically significant differences between groups either (Appendix E).

To probe further in search of at least some quantitative impact of the treatment, one more type of statistics, namely the effect size, was applied. According to Larson-Hall (2010), effect size can be “more valuable than the question of whether a statistical test is ‘significant’ or not” (p. 114). After applying the formula for calculating the effect size $r = Z/\sqrt{N}$ (Larson-Hall, 2010, p. 382), an effect size of $r = 0.23$ was found between the pre- and post-test essays of the treatment group¹. Larson-Hall (2010) considers this to be a small effect size for the groups of 100 participants (p. 266). However, for a group of 11 participants, this can be a “valuable piece of information” (p. 114), although the difference between the groups was not found to be statistically significant. Given that the writing skills were not taught explicitly, and the entire treatment period was short (re: Section 5.3 Limitations), this level of effect size should not be neglected.

To sum up the quantitative data analysis and to answer the first research question of this study, one can conclude that explicit instruction in argumentation did not impact the students’ writing skills to the extent of statistically significant changes in the treatment group’s pre- and post-test essay scores. However, explicit instruction in argumentation had an impact on the student’s writing skills to the extent of a small effect size of .23.

4.2 Qualitative data analysis

This study could have ended with the quantitative analysis presented in the previous section, having obtained the modest but unequivocal results of the quantitative impact of the treatment. However, the study posed more questions than it gave answers to. The new questions were mostly about why the treatment did not have the anticipated results, and what did not work. These questions are discussed later in this paper. Besides, the amount of effort on the teacher’s and the researcher’s part was too generous to admit to its absolute uselessness for the treatment group. These were the reasons to probe further and analyze the

¹ The corresponding values for the formula of effect size are in the SPSS output for the Wilcoxon signed rank test run to compare the pre- and post-tests of the treatment group (Appendix E).

qualitative data in search of at least some aspects of the treatment that did work, albeit without statistically significant impact. For the qualitative data, we looked into (a) the pre- and post- test essays of the students; (b) the reflection journal; (c) the feedback of the students of the treatment group during the last discussion session.

In the essays, we looked for some evidence of the improved structure, the use of linking words and basic argumentation techniques. Likewise, in the researcher's reflection journal, there were certain episodes observed during the in-class debates that could qualify as evidence of some impact of the treatment on how the students shaped their arguments during the discussions. Finally, the students' feedback on the entire treatment course gave some food for thought about what they learned and, more importantly, what they did not learn and why. These pieces of qualitative impact of the treatment, as well as the quantitative results, are discussed below.

CHAPTER FIVE: DISCUSSION AND CONCLUSION

The chapter discusses the results of the study in terms of their impact on the writing skills of the treatment group, particularly, the possible reasons for the lack of quantitative impact of the treatment.

5.1 Quantitative Impact of the Treatment: Factors that Did Not Work

As was shown in the previous section, the quantitative data analysis showed that explicit instruction in argumentation did not have statistically significant impact on the students' writing skills. However, there was some impact – to the extent of a small effect size of .23. This section discusses the possible reasons of why the treatment did not result in more tangible quantitative statically significant differences between the groups.

Various aspects of the treatment that did not have the desirable quantitative impact on the learners' writing skills and led to the lack of statistically significant impact of the treatment became evident already during but mainly after the treatment. The factors that did not work can be grouped along the following possible reasons.

One of the reasons was a virtual absence of the academic debate culture in Armenia conjured images of fight, animosity and conflict in the students. The last session in the series of debates was organized in the form of a discussion when the students were asked for positive and negative feedback on the debates that they had participated in (Appendix G).As the students mentioned during that feedback session, these images came from what they usually saw on mass media – television, the Internet, and the social networks. This cultural factor would have had less negative impact on the students had the sessions been introduced to them as “discussions” instead of “debates” from the beginning of the term.

Another reason that could be derived from observations of the debates was that the students took criticism personally. Because the students were teenagers, they were especially sensitive to criticism and disagreements during the debates. It would take time before the

students could get used to the idea that disagreements are an important feature of the formal academic debate and constitute its inalienable part. Besides, the students did not like the idea of appointing the judges from among them. The judges were appointed to involve the whole class in the discussion sessions, drawing on the study by conducted by Dundes (2001).

However, personal likes and dislikes affected their judgments. The students were not opposed to the format *per se* when the teacher was the judge. That some of the students were not comfortable with the academic debate format of the discussions became evident only during the last discussion when the students gave feedback on the entire series of the debates. During the regular discussions of the topics, the voices of those who liked the debates were more dominant. Quick dialogues immediately after the debate sessions typically conveyed positive feelings toward the debates. “Are we going to have a debated next time?” – “No” – “That’s a pity”. However, as it turned out, some of the students harbored negative feelings toward each other, and especially toward the judges, after the debates. They accused the judges of being biased, e.g., girl judges favored girls, and boy judges boys. These and other nuances, too numerous to mention here, transpired only after the very last discussion, i.e. the feedback debate. Had the researcher anticipated them before or at least early in the term, the results would have been more toward the bright side of the story.

Still another possible reason of why the treatment did not have a strong impact on the students’ writing skills was the lack of preparation before the debates. Homework and some preparation and research prior to the actual discussion of a topic are essential for the academic debate format. The students, however, needed more time to get used to the idea of a special kind of homework, i.e., research, before every in-class discussion session. They were even more reluctant to do their homework research after they had discovered that the parallel group did not have the same series of debates.

To continue the list of reasons comes the fact that the treatment was not a dedicated course but was embedded into the students' main course. It soon became evident that the students took the exercises and the debates/discussions as secondary to their main syllabus, especially when they figured out that the parallel class did not have these additional sessions. Some of the students felt privileged, but others saw it as an additional burden imposed on them, compared to the parallel group.

That the students were not told the real purpose of the debates seems to be yet another reason of the weak impact of the treatment. Although the research questions of the study are about possible relationships between teaching argumentation and the improvement of the students' writing skills, writing was not mentioned to the students during the entire treatment. It is possible that had the students been told that the goal of the debates was the improvement of their essay writing skills, they would have put more effort in their post-treatment essays.

The last, but not least, in the reasons of weak impact is the researcher's high expectations about the desirable quantitative outcomes of the treatment. Had it not been for the raised expectations, the important things that did work (re: the next section) would have been better appreciated. Because teaching argumentation enhances general learning (Lin & Mintzes, 2010) as well as language learning abilities (Reznitskaya et al., 2007), the assumption was that it would automatically result in one particular skill – writing. However, exactly because of the same reason, i.e., that teaching argumentation has such profound effects on the learner, its results cannot happen overnight, nor did they consolidate in this study during its eight weeks of treatment. This latter consideration had escaped the researcher at the beginning. But this was also a lesson learned. The present research was not about changing the students' attitudes, opinions or likes and dislikes which can indeed happen, and do happen, overnight even with adults, let alone teenagers. The ambition of this study was much greater – to tap into and change the very structure of the students' way of thinking. It is

this change that did not happen during the treatment, nor could it happen, in principle. Had the researcher understood this from the beginning, her expectations would have been more realistic.

Some of the factors listed above, e.g., the cultural factor, can be called researcher-overlooked factors. Other factors, such as the imbedded nature of the treatment, are objective limitations of the present research. There were, however, other factors that did have an impact, although their influence, as the next section shows, was not consolidated enough to translate into statistically significant quantitative results.

5.2 Qualitative Impact of the Treatment: Factors that Worked

Having established the lack of statistically significant quantitative changes resulting from the treatment, the study went on to look for at least some evidence of impact of the explicit instruction in argumentation on the writing skills of the participants. For this, the study used three sources: the essays, the researcher's reflection journal (Appendix G) and the feedback of the students during the last discussion session of the treatment period. These bits and pieces of the treatment that appeared to have some positive impact on the students' writing skills are silver linings behind the clouds that gathered on the horizon after the results of the quantitative analysis. At the same time, these tiny pieces of impact of the treatment that worked present the qualitative data of this study.

5.2.1 Evidence from the Essays

The study analyzed the essays from the following points of view:

- The lengths of the essays, i.e., the number of words
- The structure of the essays, i.e., the presence of an introduction, some supporting points and a conclusion

- The explicit use of the linking words, i.e., basic argument markers (reason, contrast and conclusion markers), as well as guarding and discounting phrases taught during the treatment

The post-test essays became longer in both comparison and the treatment groups.

However, the students of the treatment group used more linking words in their post-treatment essays compared with all the essays of the comparison group and, most importantly, compared with their own pre-treatment essays (Table 4).

Table 4. *Lengths of the essays and the use of linking words*

n/n	Comparison group				Treatment group			
	Pre-test		Post-test		Pre-test		Post-test	
	Words	Linking words	Words	Linking words	Words	Linking words	Words	Linking words
1	87	1	112	2	158	1	120	5
2	74	2	106	0	86	2	151	4
3	120	1	154	4	222	6	250	4
4	116	0	144	3	95	1	164	7
5	131	3	187	2	131	4	115	8
6	119	2	106	3	124	1	85	3
7	147	4	203	4	200	2	107	3
8	118	2	121	2	153	0	186	0
9	228	3	172	2	130	4	146	2
10	114	3	155	1	190	4	188	4
11	148	2	196	5	72	2	137	2
12	95	4	136	6				
13	139	3	143	2				
14	129	0	113	4				
15	145	0	57	1				
Average	129	2.0	140	2.7	142	2.5	150	3.8

The linking words that were used more often in the essays were “but” and “so”, respectively, as a contrast marker and a conclusion marker in argumentation. One instance of using the linking word “however” to express contrast occurred in one of the post-treatment essays of the treatment group. This structure was explicitly taught; however, it did not occur more than once, unlike “but” or “so” linking words that occurred more frequently in both

groups, probably because they are more common in everyday speech also in the students' L1. Another linking word that was used successfully was "because" in both groups, but slightly more often in the treatment group. Likewise, guarding and discounting phrases, with the use of such linking words as "maybe", "some", "many" were used in both groups, but slightly more so in the treatment group.

Two instances in the treatment group of looking at the topic of the essay from two opposite perspectives also happened, with the use of the linking word "but". For example, after a paragraph on the use of smart phones by everyone, followed the sentence, "But wouldn't it be a disaster?" Because debates exposed the participants to more than one point of view on the topics discussed, this episode can be viewed as a direct result of the topic "Are videogames good for kids?" debated during the treatment.

Regarding the use of introductions and conclusions, with or without the use of conclusion markers, was almost similar for both groups. However, the treatment group used more explicit conclusion markers, such as the words "so" and "conclusion".

Given that the groups' regular textbook and the teaching syllabus did not contain the essay composition structure, it seems legitimate to attribute the emergence/ occurrence of even the smallest signs of a structure in the students' writings to the treatment or at least to that innate "argument schema" (Reznitskaya et al., 2007, p. 450), which the treatment activated, or "modified" in the authors' parlance – not entirely unlike the innate linguistic capacity in Chomski's (1980) theory or Universal Grammar (UG). The existence of the argument schema was only briefly mentioned in the literature review; here, however, in the context of the study's treatment syllabus, a closer look seems in order. As mentioned by Reznitskaya et al. (2007), "knowledge consists of generic mental structures, or schemas" and "learning involves modification of these schemas". For such modifications to occur, the authors' teaching practices were "group discussions of controversial issues and explicit

instruction in principles of argumentation” (p. 450), which is exactly what the treatment syllabus of this study attempted (re: Appendix B).

5.2.2 Evidence from Observed Episodes during In-class Debates

Two noteworthy episodes occurred during the heated discussion in one of the debate sessions whose topic was “Are videogames good for kids?” This topic was the students’ favorite. By their request, it continued also in the next debate. One student from the negative team used a sequence of guarding phrases in her argument against videogames, as she went from her original unguarded statement “Videogames bring nothing but violence” to a more guarded claim “Some violence exists in all videogames” to a perfectly guarded argument “Some violence exists in most videogames”. The students of both teams, together with the teacher, observed the effect of this common move in argumentation, when the affirmative team had no other choice but to agree with this argument of the negative team. Had the researcher had a chance to intervene, it would be a perfect example on which to explain this technique of argumentation, in addition to the written exercises that the students did during the lessons.

The described episode testifies to the fact that there is an innate “basic argumentation schema” (Reznitskaya et al, 2007, p. 455) discussed in the previous section. This student was also the most advanced in the group, which is in agreement with the teachability and learnability components of Manfred Pienemann’s (2015) Processability Theory discussed in the literature review of this paper. In other words, the interlanguage of this student was ready to process the guarding technique included in the treatment syllabus.

The second noteworthy episode happened in the heat of the same debate, when one of the students used a perfect discounting phrase: “Although the Affirmative team says that videogames are good for kids, there is evidence that they bring more violence”.

One more episode happened during the unique debate whose setup was Students vs. Teachers. The topic, “Cars or Bicycles” as well as the setting were suggested by students themselves (re: Appendix B). Someone said: “So what, your conclusion is missing!” about an argument made by another student. Students start paying attention to the structure of their speech in general – not necessarily during the formal debates.

These seeds of results, however, needed more nourishment before they could sprout and become strong enough to lend themselves to quantification and more substantial analysis. But, humble as the above results are, they are silver linings behind the clouds of this research. Given more time, they are bound to grow stronger to form patterns that will be more systematic in the students’ writings. Provided, that is, that the course is more than the mere 16 to 18 hours of teaching and is dedicated instead of being merely annexed to the core course.

5.2.3 Evidence from the Students’ Feedback

The students appreciated the noticeable enrichment of their vocabulary due to the routine exposure to the debate discourse and its technical terms, such as the names of the teams, of the construction speeches, the topics, and other terms. The students acquired this vocabulary implicitly while concentrating on the meaning of the debated topics. In other words, they learned the new vocabulary in analogy with “focus-on-form” approach to learning that “draws learners’ attention to a linguistic form while they are primarily focused on meaning – i.e. trying to communicate” (Ellis, 2015, p. 242).

It is important to mention in this respect that the topic-related vocabulary was acquired through research, googling of the upcoming topics and the delivery by the peers during the debates. The academic debate-related vocabulary was learned implicitly, through multiple exposures to the same debate terminology in the contexts of the various debate

topics. The effect was reinforced by the fact that most of the topics were suggested by the students themselves, i.e., were most interesting and engaging for them.

Casting a retrospective glance at these precious pieces of evidence, one is almost grateful that the quantitative analysis returned “no results”. For, if the answer to the first research question were “yes”, i.e., if the quantitative analysis confirmed the existence of a statistically significant relationship between teaching of argumentation and the improvement of writing skills, no subsequent qualitative analysis would have ensued. In other words, these pieces of evidence would be “ironed out” as an inevitable price paid to the triumphant results of the quantitative research.

These pieces of evidence of a certain positive impact of the treatment that come from the students’ essays, from the observed episodes of in-class debates and from the students’ feedback do not undermine the results of the qualitative analysis that returned no significant changes between the groups after the treatment. They, however, give some hope. The researcher’s belief in argumentation and debate as an approach to teaching and learning a language is too firm to be shattered by anything, much less by a single event such as, for example, the discovery that the quantitative impact of the treatment was negligible. This Capstone project is an opportunity for the researcher to learn how to teach argumentation and academic debate and, as such, is the beginning of a long journey – a journey that is, by definition, all about the process and not the results.

This Capstone project, therefore, was process-oriented. Of course, the rejection of the Null Hypothesis would have been more desirable. However, even though the null hypothesis is confirmed, it can be reasonably explained by many factors, e.g., by limitations, such as the time constraints, the researcher’s own lack of experience, the arrangement that the teaching was not direct but was done through the proxy of the permanent teacher. These and other limitations addressed in the next section, are enough not to attribute the weak quantitative

impact to the method itself. On the contrary, it rather means that it is worthwhile to pursue the proposed methodology further. So this Capstone project is the beginning of a long journey.

5.3 Limitations

The nonexistent culture of academic debates in Armenia was a limiting factor. The students perceived the debates as competitions, even confrontations or fighting grounds, and an opportunity to establish a truth, rather than intellectual exercises. No matter how much the teacher emphasizes the collaborative and intellectual nature of the debates, much more time is needed for these messages to take root. To mitigate the effect of this limitation, the teacher took upon herself the responsibility of moderating the debates and managing the disagreements. As was mentioned in the literature review, debate moderation can also be done by the students themselves (Horn, 2008). However, in the present research it was done by the teacher because of the different cultural background, the culture of the academic debate being virtually non-existent in Armenia.

Connected with this is another limitation, namely the students' reluctance in doing homework research in preparation for the structured academic debates.

The teaching of argumentation in the treatment group was annexed to the regular syllabus of the program, which resulted in additional learning burden for the students of that group. To compensate for this extra work on the students' part, the teacher had to devote a commensurate effort and time to keep the students motivated throughout the entire teaching period. As a result, the amount of time available for the implementation of the treatment syllabus was limited to a maximum of 20 teaching hours.

To maintain the purity of the experiment and to avoid the Hawthorn effect (Dornyei, 2007), the students of both groups, and more importantly, of the treatment group, were not told about the experiment. Before long, however, they learned the fact that the parallel group

did not have the additional exercises and series of debates. They asked the teacher and the researcher as to why. Having received an ambiguous answer that was only half the truth, at best, they, as it is often the case with teenagers, appeared less excited about the debates. This limitation is an affective factor related to Krashen's (1982) Affective Filter hypothesis about "affective variables acting to impede or facilitate" learning (p. 32). So, this particular limitation acted to impede learning.

Another limitation was the small sample size – 16 and 11 students in the comparison and the treatment group, respectively. Moreover, the participants of both groups and the groups themselves were not selected randomly but from among the available options, which resulted in the convenience/opportunity sampling, "the most common sample type in L2 research" (Dornyei, 2007, p. 98). Still another limitation was the time/duration of the course.

The lack of experience in designing argumentation-related tasks (on the part of the researcher) was also a limitation. The outcome of the experiment with no statistically significant differences between the two groups (i.e. when the null hypothesis is confirmed) was non-conclusive; it is theoretically and at least partially attributable to this limiting factor and not to the actual/positive absence of influence of explicit instruction in argumentation on the writing skills.

5.4 Delimitations

The results of this study can only have meaningful interpretations when applied to the participants or to other students of the same L2 proficiency level (low elementary) and approximately the same age (12 – 16 years) coming from the same or similar language schools. The way of teaching the basics of argumentation, i.e., the three basic elements of argumentation, also delimits the results. However, since by all evidence obtained so far, this is the first such attempt in Armenia, other similar schools in the country may benefit from familiarizing themselves with the methodological approach and the discussion of the results.

5.5 Recommendations: What to Do Differently

There are several aspects of this experimental study that the researcher would do differently, if the experiment were repeated. They seem obvious after the study is over. However, some simple but important factors were overlooked during the research. Especially, this concerns the treatment component of the research.

The study will benefit if the researcher reads the pre-treatment essays of both groups as early as possible – ideally, immediately after the students write them. Equipped with the knowledge of students' strengths and weaknesses, particularly with respect to the patterns of their use of linking words in writing, the researcher can develop a more efficient treatment syllabus. For example, the in-class exercises for teaching the linking words will be based on the students' writing patterns, i.e., tailored to the students' needs. Whereas, in the present study all essays, both pre- and post-treatment, were read after the treatment stage was already over. Late reading of the essays was not, however, without reason, one of them being that the selection of the treatment group was unbiased and based solely on the number of the students. The smaller group was to become the treatment group, on the grounds that smaller groups of learners provide better opportunities for all students to be engaged and, therefore, are better settings for structured debates. In addition, smaller groups (around 10 students) do not limit the researcher's ability to observe the group during the discussion (Dundes, 2001).

Another aspect of the study to be done differently concerns the students' feedback. It is important to solicit the student's feedback at least twice during the treatment, one in the middle and the other in the end. This will give the researcher the opportunity and enough information to adjust the treatment syllabus accordingly. As it were, however, in the present study, students gave feedback on their perceptions, attitudes and feeling about the debates only toward the end of the treatment. One more consideration about different ways of

conducting the same study is not to label the oral discussions “debates” but to call them “discussions” instead. This issue was discussed earlier in the paper.

They say, the best way of learning something is to teach it to others. This study has only scratched the surface of the field of argumentation and debate, let alone its impact on learners’ L2. Counterintuitive though it may sound, it was not easy to teach the surface, i.e., the basics only. Indeed, it was even more difficult since the researcher herself still did not know many important things that lay at deeper layers. Irrespective of what could be done differently, these are also lessons learned and, as such, are among the most valuable outcomes of the unique experience of this study.

Thus ends the bittersweet story of one Capstone thesis project in TEFL.

References

- Brown, N. (2009). Argumentation and debate in foreign language instruction: A case for the traditional classroom facilitating advanced-level language uptake. *The Modern Language Journal*, 93(4), 534-549. Retrieved from <http://www.jstor.org/stable/25612230>
- Chin, C., & Osborne, J. (2010). Supporting Argumentation Through Students' Questions: Case Studies in Science Classrooms. *The Journal of the Learning Sciences*, 19(2), 230-284. Retrieved from <http://www.jstor.org/stable/27801117>
- Chomsky, N. (1980). *Rules and representations*. New York: Columbia University Press.
- Conner, A., Singletary, L., Smith, R., Wagner, P., & Francisco, R. (2014). Teacher support for collective argumentation: A framework for examining how teachers support students' engagement in mathematical activities. *Educational Studies in Mathematics*, 86(3), 401-429. Retrieved from <http://www.jstor.org/stable/43589858>
- Dawson, V., & Venville, G. (2010). Teaching strategies for developing students' argumentation skills about socioscientific issues in high school Genetics. *Research in Science Education*, 40(2), 133-148. doi: 10.1007/s11165-008-9104-y
- De Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition*, 10, 7–21.
- Dornyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.
- Dundes, L. (2001). Small Group Debates: Fostering Critical Thinking in Oral Presentations with Maximal Class Involvement. *Teaching Sociology*, 29(2), 237-243. Retrieved from <http://www.jstor.org/stable/1318721>
- Ellis, R. (2015). *Understanding second language acquisition* (2nd edition). Oxford: Oxford University Press.

- Hickey, P., & Watson, T. (2014). *Lingua Anglia: Bridging Language and Learners: The English Journal*, 104(2), 121-123. Retrieved from <http://www.jstor.org/stable/24484432>
- Horn, I. (2008). Accountable argumentation as a participation structure to support learning through disagreement. *Journal for Research in Mathematics Education. Monograph*, 14, 97-126. Retrieved from <http://www.jstor.org/stable/30037742>
- Hughes, A. (1989). *Testing for Language Teachers*. Ch. 5. Cambridge University Press, UK.
- Johns, A. (1993). Written Argumentation for Real Audiences: Suggestions for Teacher Research and Classroom Practice. *TESOL Quarterly*, 27(1), 75-90.
doi:10.2307/3586952
- Kim, J., & Lee, H. (2017). Enhancing student key competencies through socioscientific issues instruction. Korean Association for Learner-Centered Curriculum and Instruction (abstract)
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Pergamon Press Inc.
- Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York: Routledge
- Learning through arguing: The keys to successful classroom debates.
<http://busyteacher.org/10319-learning-through-arguing-the-keys-to-successful.html>
- Lin, S., & Mintzes, J. (2010). Learning argumentation skills through instruction in socioscientific issues: The effect of ability level. *International Journal of Science and Mathematics Education*, 8(6), 993-1017.
- Marttunen, M. (1997). Electronic Mail as a Pedagogical Delivery System: An Analysis of the Learning of Argumentation. *Research in Higher Education*, 38(3), 345-363. Retrieved from <http://www.jstor.org/stable/40196300>

- Morgan, A. (2006). Argumentation, Geography Education and ICT. *Geography*, 91(2), 126-140. Retrieved from <http://www.jstor.org/stable/41555404>
- Morse, K. (2013). Debate: A tool for language learning. *Journal of Border Educational Research*, 10(1).
- Moss, L. (2011). Syllogistic Logic with Comparative Adjectives. *Journal of Logic, Language, and Information*, 20(3), 397-417. Retrieved from <http://www.jstor.org/stable/41488486>
- Painter, C., Coffin, C., & Hewing, A. (2003). Impacts of directed tutorial activities in computer conferencing: a case study. *Distance Education*, 24(2), 159-173. doi:10.1080/0158792032000127455
- Pienemann, M. (1998). *Language processing and second language development: Processability theory*. Amsterdam: Benjamins.
- Pienemann, M. (2015). An outline of processability theory and its relationship to other approaches to SLA. *Language Learning*, 65(1), 123-151. doi:10.1111/lang.12095.
- Pienemann, M., & Håkansson, G. (2007). Full transfer vs. developmentally moderated transfer: A reply to Bohnacker. *Second Language Research* 23(4), 485-493.
- Rapanta, C., Garcia-Mila, M., & Gilabert, S. (2013). What Is Meant by Argumentative Competence? An Integrative Review of Methods of Analysis and Assessment in Education. *Review of Educational Research*, 83(4), 483-520. Retrieved from <http://www.jstor.org/stable/24434220>
- Reznitskaya, A., Anderson, R., & Kuo, L. (2007). Teaching and Learning Argumentation. *The Elementary School Journal*, 107(5), 449-472. doi:10.1086/518623.
- Rybold, G. (2006). *Speaking, listening and understanding: Debate for non-native English speakers*. New York, NY: International Debate Education Association.

Sinnott-Armstrong, W., & Neta, R. (2011). *“Think again: How to understand arguments”*.

An online course of Duke University. <https://www.coursera.org/learn/understanding-arguments>

Song, P., Graf, E., & Rijn, P. (2013). Using argumentation learning progressions to support teaching and assessments of English language arts. *R&D Connections*, 22, 1-14.

Toulmin, S. (2003). *The uses of argument*. Cambridge: Cambridge University Press.

Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Weir, C. (1993). *Understanding and developing language tests*. Prentice Hall International, UK.

Zhang, Y. (2005). Processing and formal instruction in the L2 acquisition of five Chinese grammatical morphemes. In Pienemann, M. (ed.), *Cross-linguistic aspects of processability theory*, 155-178. Amsterdam: John Benjamins.

Zhang, Y., & Wang, J. (2012). The elaboration of cultivating learners' English communicative competence in China. *English Language Teaching*, 5(12), 111.

Appendices

Appendix A. Scoring Guide

Score	Description
5	<ul style="list-style-type: none">- The composition clearly demonstrates competence in writing- is well organized and developed- effectively addresses the task- displays good knowledge of grammar and sentence structure
4	<ul style="list-style-type: none">- The composition demonstrates competence in writing- in general is well organized and developed, though it may have some drawbacks- addresses some parts of the task more effectively than others- displays general knowledge of grammar and sentence structure, though there may be errors
3	<ul style="list-style-type: none">- The composition demonstrates minimal competence in writing- may be adequately addressed but not sufficiently developed- addresses the topic but may slight parts of the task- may contain some serious errors that occasionally obscure meaning
2	<ul style="list-style-type: none">-The composition suggests incompetence in writing- inadequate organization and development- fails to address the topic appropriately- an accumulation of errors in sentence structure/grammar
1	<ul style="list-style-type: none">- The composition demonstrates incompetence in writing, contains serious and persistent writing errors, may be illogical, or may reveal the writer's inability to comprehend the question.

Appendix B. Treatment Syllabus for the instruction in argumentation skills in the treatment group (18 hours)¹

n/n	Date	Debate/Discussion topic	Grammar / Argumentation techniques	Speaking / Format	Useful links for homework research before discussion
1	Jan.10 & 12, 2018	To lie or not to lie?	Getting a taste for debating Setting a spirit of questioning and healthy skepticism	1. Introduction: state your position in the debate - "affirmative" or "negative" 2. At least three reasons with supporting details for your position (you can have more than three reasons, but make sure that they do not repeat one another) 3. Conclusion: re-state your position and your reasons (in one or two sentences) - for everyone to know that you've done talking.	https://debate.bard.edu/?page_id=2221 https://www.npr.org/2016/12/27/506314053/after-half-a-century-inmates-resurrect-the-norfolk-prison-debating-society Harvard debate team loses to NY prisoners: https://www.theguardian.com/education/2015/oct/07/harvards-prestigious-debate-team-loses-to-new-york-prison-inmates
2	Jan. 16 & 18, 2018	Should cell phones be allowed in the classroom?	Linking words of reason, (reason markers): because, since, for, as	Two teams of four students (flip of a coin) Judge panel of three students	http://www.debate.org/opinions/should-cell-phones-be-allowed-in-school
3	Jan. 20& 23, 2018	Are videogames good for kid?	Discounting technique	Two teams of four students (flip of a coin) Judge: the teacher	http://www.debate.org/opinions/are-video-games-good-for-kids?nsort=5&ysort=2
4	Jan. 25 &30, 2018	Are videogames good for kids? (continued) Student-suggested	Linking words of contrast: although, but, however, despite. Paraphrase exercise 1: Paraphrase “used to do something” using contrast	Two teams of four (students decide) No judges	http://www.debate.org/opinions/are-video-games-good-for-kids?nsort=5&ysort=2 http://www.debate.org/opinions/are-video-games-bad-for-you

¹ The vocabulary will be taught implicitly, together with speaking and argumentation skills. In this chart it has been integrated in the activities of the corresponding columns.

			markers (Appendix B-1)		
5	Feb. 1 2018	Is it always good to make jokes? Student-suggested	-	Two teams of four (students decide) Judge: the teacher	s://www.helpguide.org/articles/relationships-communication/managing-conflicts-with-humor.htm https://www.quora.com/How-does-Chinese-humor-differ-from-American-humor
6	Feb. 6 & 8, 2018	Time-travelling (From English in Mind)	Guarding technique	Two teams of four students (flip of a coin) Judge panel of three students	https://www.quora.com/Would-you-rather-travel-10-years-into-the-past-or-10-years-into-the-future-Why https://waitbutwhy.com/table/time-machine-question
7	Feb. 13 &15, 2018	Study with or without homework? Student-suggested	Conclusion markers: so, therefore	Two teams of four students (flip of a coin) Judge: the teacher	http://time.com/4466390/homework-debate-research/
8	Feb. 20 &22, 2018	Study with books or digital technology? Student-suggested	Paraphrase exercise 2: Paraphrase the 3 rd conditional using reason and conclusion markers (Appendix B-2)	Two teams of four (students decide) No judges	
9	Feb. 27, 2018	Using cars vs. bicycles? Student-suggested	-	Two teams: Students vs. Teachers	http://top10hell.com/top-10-countries-with-most-bicycles-per-capita/ Happiest and wealthiest countries: https://www.nationalgeographic.com/travel/top-10/2017-worlds-happiest-countries/ https://www.worldatlas.com/articles/the-richest-countries-in-the-world.html Health benefits: https://www.betterhealth.vic.gov.au/health/healthyliving/cycling-health-benefits
10	Mar. 1, 2018	Pros and cons of the debate series: Students' feedback	-	Free discussion: two teams of students	

Appendix B-1: Paraphrase exercise 1

Adapted from: <http://www.perfect-english-grammar.com/used-to.html>
<http://www.perfect-english-grammar.com/used-to-exercise-1.html>

Used To Do Something (explanation)

'Used to + infinitive':

We use this expression to talk about habits or repeated actions in the past which we **don't do in the present**. We also use it to talk about states in the past which are **no longer true**.

Paraphrase Exercise

Instruction: Paraphrase the sentences about the past (with “used to + infinitive”). Use the linking words of contrast (but, however, despite, although) to say something about the present.

Example:

1) She used to love chocolate.

Paraphrase: She used to love chocolate, but now she loves lollipops.

2) They used to live in India.

Paraphrase: _____

3) We used to go to the beach every summer.

Paraphrase: _____

4) He used to smoke.

Paraphrase: _____

5) I used to play tennis when I was at school.

Paraphrase: _____

6) She used to be able to speak French.

Paraphrase: _____

7) He used to play golf every weekend.

Paraphrase: _____

8) They both used to have short hair.

Paraphrase: _____

9) Julie used to study Portuguese.

Paraphrase: _____

10) I used to hate school.

Paraphrase: _____

Appendix B-2: Paraphrase exercise 2

Adapted from: <http://www.perfect-english-grammar.com/third-conditional.html>

The Third Conditional (explanation)

We make the third conditional by using the **past perfect** after 'if' and then 'would have' and the **past participle** in the second part of the sentence:

- if + past perfect, ...would + have + past participle

It talks about the past. It's used to describe a situation that didn't happen, and to imagine the result of this situation.

- If she **had studied**, she **would have passed** the exam (but, really we know she didn't study and so she didn't pass)
- If I **hadn't eaten** so much, I **wouldn't have felt** sick (but I did eat a lot, and so I did feel sick).
- If we **had taken** a taxi, we **wouldn't have missed** the plane
- She **wouldn't have been** tired if she **had gone** to bed earlier
- She **would have become** a teacher if she **had gone** to university
- He **would have been** on time for the interview if he **had left** the house at nine

Paraphrase Exercise

Instruction: Fill in the gaps to make the third conditional.

Paraphrase the sentences using reason markers (**because, since, as**) and result/conclusion markers (**so, therefore, consequently, accordingly, as a result**).

Example:

1) If she _____ (study) she _____ (pass) the exam.

3rd conditional: If she had studied, she would have passed the exam.

Paraphrase (option 1): **Because** she did not study, she did not pass the exam.

Paraphrase (option 2): She did not study; **so** she did not pass the exam.

2) If you _____ (not/be) late, we _____ (not/miss) the bus.

3rd conditional: _____

Paraphrase: _____

3) If we _____ (arrive) earlier, we _____ (see) John.

3rd conditional: _____

Paraphrase: _____

4) If they _____ (go) to bed early, they _____ (not/wake) up late.

3rd conditional: _____

Paraphrase: _____

5) If he _____ (become) a musician, he _____ (record) a CD.

3rd conditional: _____

Paraphrase: _____

6) If she _____ (go) to art school, she _____ (become) a painter.

3rd conditional: _____

Paraphrase: _____

7) If I _____ (be) born in a different country, I _____ (learn) to speak a different language.

3rd conditional: _____

Paraphrase: _____

8) If she _____ (go) to university, she _____ (study) French.

3rd conditional: _____

Paraphrase: _____

9) If we _____ (not/go) to the party, we _____ (not/meet) them.

3rd conditional: _____

Paraphrase: _____

10) If he _____ (not/take) the job, he _____ (go) travelling.

3rd conditional: _____

Paraphrase: _____

11) He _____ (be) happier if he _____ (stay) at home.

3rd conditional: _____

Paraphrase: _____

12) She _____ (pass) the exam if she _____ (study) harder.

3rd conditional: _____

Paraphrase: _____

Appendix C. Scores on pre- and post-treatment essays and descriptive statistics

n/n	Comparison group		Treatment group	
	Pre-treatment	Post-treatment	Pre-treatment	Post-treatment
1	3.0	2.8	2.5	3.5
2	3.0	1.8	4.0	4.3
3	3.0	3.3	4.5	4.0
4	3.8	4.0	3.5	3.3
5	3.3	3.3	4.5	3.3
6	3.0	2.8	3.3	2.5
7	4.5	4.8	3.5	3.0
8	3.0	3.3	3.5	3.0
9	3.3	3.0	3.3	3.3
10	2.8	4.5	4.25	4.8
11	4.3	4.0	2.8	3.0
12	4.0	2.8		
13	3.3	3.3		
14	3.5	3.5		
15	3.3	2.3		
Mean	3.4	3.3	3.6	3.4
Mode	3.0	3.25	3.5	3.25
Median	3.3	3.3	3.5	3.3

Appendix D. Parametric tests results (Independent-samples t-tests and Paired-samples t-tests)

Independent samples t-test: to compare the means of the pre-test scores of the two groups

Group Statistics

Group_code	N	Mean	Std. Deviation	Std. Error Mean
C_T_group_Pre Comparison group	15	3.407	.5175	.1336
Treatment group	11	3.605	.6529	.1968

Independent Samples Test

	Levene's Test for Equality of Variances		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
C_T_group_Pre Equal variances assumed	.661	.424	-.863	24	.397	-.1979	.2294	-.6712	.2755
Equal variances not assumed			-.832	18.529	.416	-.1979	.2379	-.6967	.3009

Independent samples t-test: to compare the means of the post-test scores of the two groups

Group Statistics

Group_code		N	Mean	Std. Deviation	Std. Error Mean
C_T_group_Post	Comparison group	15	3.300	.7928	.2047
	Treatment group	11	3.455	.6654	.2006

Independent Samples Test

	Levene's Test for Equality of Variances		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	
C_T_group_Post	Equal variances assumed	.127	.725	-.524	24	.605	-.1545	.2947	-.7628	.4537
	Equal variances not assumed			-.539	23.482	.595	-.1545	.2866	-.7468	.4377

Paired sample t-test: to compare the means of the pre- and post- tests of the comparison group

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 Comp_Pre	3.407	15	.5175	.1336
Comp_Post	3.300	15	.7928	.2047

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 Comp_Pre - Comp_Post	.1067	.7156	.1848	-.2896	.5030	.577	14	.573

Paired sample t-test: to compare the means of the pre- and post- tests of the treatment group

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Treatm_Pre	3.605	11	.6529	.1968
	Treatm_Post	3.455	11	.6654	.2006

Paired Samples Test

		Paired Differences				t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower				Upper
Pair 1	Treatm_Pre - Treatm_Post	.1500	.6368	.1920	-.2778	.5778	.781	10	.453

Appendix E. Non-parametric tests results (Mann-Whitney U tests and Wilcoxon signed-rank tests)

Mann-Whitney Test: to compare the means of the pre-test scores of the two groups

Ranks

Group_code		N	Mean Rank	Sum of Ranks
C_T_group_Pre	Comparison group	15	12.17	182.50
	Treatment group	11	15.32	168.50
Total		26		

Test Statistics^b

	C_T_group_Pre
Mann-Whitney U	62.500
Wilcoxon W	182.500
Z	-1.051
Asymp. Sig. (2-tailed)	.293
Exact Sig. [2*(1-tailed Sig.)]	.305 ^a

a. Not corrected for ties.

b. Grouping Variable: Group_code

Mann-Whitney Test to compare the means of the post-test scores of the two groups

Ranks

Group_code		N	Mean Rank	Sum of Ranks
C_T_group_Post	Comparison group	15	12.83	192.50
	Treatment group	11	14.41	158.50
Total		26		

Test Statistics^b

	C_T_group_Post
Mann-Whitney U	72.500
Wilcoxon W	192.500
Z	-.526
Asymp. Sig. (2-tailed)	.599
Exact Sig. [2*(1-tailed Sig.)]	.610 ^a

a. Not corrected for ties.

b. Grouping Variable: Group_code

Wilcoxon Signed-Ranks Test to compare the means of the pre- and post-test scores of the comparison group

Ranks

		N	Mean Rank	Sum of Ranks
Comp_Post - Comp_Pre	Negative Ranks	7 ^a	6.57	46.00
	Positive Ranks	5 ^b	6.40	32.00
	Ties	3 ^c		
	Total	15		

a. Comp_Post < Comp_Pre

b. Comp_Post > Comp_Pre

c. Comp_Post = Comp_Pre

Test Statistics^b

	Comp_Post - Comp_Pre
Z	-.554 ^a
Asymp. Sig. (2-tailed)	.579

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

Wilcoxon Signed-Ranks Test to compare the means of the pre- and post-test scores of the Comparison group

Ranks

		N	Mean Rank	Sum of Ranks
Treatm_Post - Treatm_Pre	Negative Ranks	6 ^a	5.75	34.50
	Positive Ranks	4 ^b	5.12	20.50
	Ties	1 ^c		
	Total	11		

a. Treatm_Post < Treatm_Pre

b. Treatm_Post > Treatm_Pre

c. Treatm_Post = Treatm_Pre

Test Statistics^b

	Treatm_Post - Treatm_Pre
Z	-.716 ^a
Asymp. Sig. (2-tailed)	.474

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

Appendix F. Cronbach's alpha test results

Case Processing Summary

		N	%
Cases	Valid	26	50.0
	Excluded ^a	26	50.0
	Total	52	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.651	.656	2

Inter-Item Correlation Matrix

	C_T_group_Pre_rater_1	C_T_group_Pre_rater_2
C_T_group_Pre_rater_1	1.000	.488
C_T_group_Pre_rater_2	.488	1.000

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
C_T_group_Pre_rater_1	2.962	.518	.488	.238	. ^a
C_T_group_Pre_rater_2	3.981	.390	.488	.238	. ^a

a. The value is negative due to a negative average covariance among items. This violates reliability model assumptions. You may want to check item codings.

Appendix G. Observation-Reflection Journal

December 14, 19 and 21, 2017

Observations of both groups

Characteristics of the two groups

Group #1: 16 students, one was absent that day

Group #2: 11 students

There was only observation of the students' regular class, a textbook English in Mind, low intermediary. No additional activities were done.

The students in both groups seem to be similar, in terms of engagement, attentiveness, question asking, or motivation. Age range is also the same.

Considerations for the selection of the treatment group: the smaller group (11 students) seems to be more appropriate, to keep all the students engaged in the upcoming debates, rotation of judges in the judge panel, etc.

January 10, 2018

The first debate in the treatment group: To lie or not to lie?

The topic was researcher-suggested. The purpose of this first debate session was to introduce the students to the structured academic debate format, so that the students get the taste for debating.

One of the students asked a question about using facts, it seemed important to him. The teacher connected this to the importance of researching the topic by at least googling it before the debate. The student and the group seemed to understand the point.

It is important to explain the difference between fact and opinion, for example:

- “they become stupid” – a “weak” point, i.e., a point that is very easy to detect, argue against and gain a point. It is an opinion, and you do not argue against this opinion, you do not argue for this opinion, but against the fact that it is not supported by evidence or by facts

The students seemed very open and receptive to the new ideas, e.g., to the new grading template/rubric for the judges.

The first debate was not very well structured. Although the grading rubric emphasized “respect for the opposite team” as one of the grading criteria, the students kept interrupting one another. This was anticipated before the debate. On the whole, the students seemed to like the new experience.

January 12, 2018

No-debate day and only introducing the basic structure of the constructive speeches of the two parties to the debate, affirmative and negative

1. Introduction: state your position in the debate - "affirmative" or "negative"

2. At least three reasons with supporting details for your position (you can have more than three reasons, but make sure that they do not repeat one another)

3. Conclusion: re-state your position and your reasons (in one or two sentences) - for everyone to know that you've done talking.

To better convey the idea, the template was presented in the format of “tell the audience what you want to tell, then tell it, then tell the audience what you’ve just told”. The students got the idea, but will they use it in their actual performance next time?

January 16, 2018

The topic is Should cellphones be allowed in the classroom?

Format: Two teams of four students (decided on the flip of a coin). The other students made a Judge panel and were given the grading rubric.

Classroom Debate Rubric for Judges

Criteria	5 points	4 points	3 points	2 points	1 point
Respect for the other team, e.g.: <ul style="list-style-type: none"> • listening attentively • not interrupting 					
Use of facts – assuring, e.g., <ul style="list-style-type: none"> • cite famous people • use of facts 					
Understanding of the topic, e.g., <ul style="list-style-type: none"> • using the opponent’s weak points • using the right terminology (evidence of homework research) 					
Timing <ul style="list-style-type: none"> • preparation time (3 min) • constructive speech (6 min) • rebuttal (3 min) • concluding speech (3 min) 					
Total points					

The primary importance of respect for the other team (the first and most important criterion in the rubric) was especially emphasized.

The students seemed to like the debate, on the whole. However, they did not like who the judges graded their performance. They suggested that the teacher be the judge next time.

Some expressions during the debate:

- “they become stupid” – a “weak” point, i.e., a point that is very easy to detect, argue against and gain a point. It is an opinion, and you do not argue against this opinion, you do not argue for this opinion, but against the fact that it is not supported by evidence or by facts.

Taking advantage of the opponent’s weak arguments earns you an additional point in your final grade.

- “All interrupted the opposite team”. Unguarded expression, guard your expressions by
Start introducing guarding terms and expressions
- A judge’s remark “Team members couldn’t introduce their minds”

Use of **liking words**

Time to introduce: reason markers exercises and construction speech’ structures

January 18, 2018

No-debate class.

The grammar and basic skills of argumentation were presented:

Linking words of reason (reason markers): because, since, for, as

It was an in-class exercise, the students are used to these types of exercises, did not seem to mind, and overall did well. About “since”: they knew the meaning of this word as a time indicator, but not as a synonym of “because”. Well, they know now, but will they use this word in the new sense?

Apparently, this is going to be not a single, on-shot experiment but the beginning of a long series of trials and errors, observations, corrections and of course collection and analysis of information about the entire process.

Students asked “Does the other group do these debates too?” they start recognizing patterns in the routine of debates and asking questions.

January 20, 2018

The debate topic: Are videogames good for kids?

The two teams of four students were decided on the flip of a coin. Taking into consideration their suggestions after the previous debate, this time the teacher was the judge. The students seemed to like this format better. But it has its shortcomings. The three students that were not in the debating team, had nothing to do and started playing with their cellphones. On the other hand, if they joined the debating teams, the teams will be too big – five and six people. This issue has to be addressed.

The topic was student-suggested.

Two noteworthy episodes happened during the debate:

1. One perfect example of using guarding techniques by going from unguarded to more guarded phrases:

“Nothing but violence” – expression used by the negative team, then it was replaced by a more guarded expression: Videogames bring nothing but violence → Some violence exists in all videogames → Some violence exists in most videogames

2. In the heat of the debate, one of the students used a discounting phrase:

“Although the Affirmative team says that videogames are good for kids, but there is evidence that they bring more violence”.

It’s time to introduce the technique of discounting in the argumentation.

It is extremely important to explain to the students that the main goal of these debates, and indeed of academic debates in general, is not to win but to participate. It’s a win-win game, etc. To win in an academic debate means not to convince those who do not agree with you (because people seldom change their convictions over a short period of time, especially about controversial and sensitive topics). To win an academic debate or an argument means to show / demonstrate the beauty of your speech and especially the beauty of your arguments.

So that even those who do not agree with you, admire your speech.

Students are getting carried away by emotions because the teams were formed voluntarily and each contained students according to his/her genuine convictions.

More debate resolutions connected with the textbook topics

January 23, 2018

Discounting technique was introduced to the students, with exercises, videos, and examples. The students understood, and even brought their own examples.

Explain the word “discounting” and bring more examples of discounting phrases

There is no better way of learning something than to teach it to others.

We only scratch the surface of the argumentation theory (still emerging in my mind), and we teach even part of this surface, i.e., the very basic things only.

Still, it’s not easy (indeed even more difficult) to teach the surface, even the very basic things, w/o knowing what’s deeper.

January 25, 2018

By the students’ request, the topic of videogames continued. There were two teams of four, voluntarily formed and the game continued without judges. There were three students absent that day, and no “idle” students were in the class. The students brought more examples from

their own experience, e.g., “eyesight getting worse”, and the discussion was as heated as the previous day.

January 30, 2018

It was the researcher’s idea use the grammatical structures already familiar to the students to teach the basic techniques of argumentation, for example, connect the familiar structure “used to do something” from the students’ regular textbook to the expressions that contain contrast markers – although, but, however, despite. For this, a paraphrase exercise was given to the students.

Paraphrase exercise #1:

“used to do smth.” – into contrast markers

There was another paraphrase exercise (re: class of February 22) that required the students to paraphrase the 3rd conditional into sentences that included reason and conclusion markers.

February 1, 2018

Is it always good to make jokes? Student-suggested

Two teams of four (students decide), the teacher is the judge.

Motivation in a debate: low, because Ss wanted to take sides according to their own beliefs

Note taking: the teacher doing it for them on the whiteboard, but

Next time:

Make ss to take notes during the debate

Appoint a moderator from among the students

Questions to ask in class: e.g., choose the next debate topic

- If you could take part in a flash mob, would you?
- Is it good for celebrities to give charity?
- Learning a foreign language: A waste of time?
- Suggest your own topic

February 6, 2018

Debate topic: Time-travelling (From English in Mind)

Two teams of four students (flip of a coin)

Judge panel of three students: the students did not seem to mind any more, perhaps because they have already got used to the format and were more comfortable with being even in the judge panel.

February 8, 2018

Guarding technique

Introducing the guarding technique of the argumentation was much easier than the discounting technique. The corresponding grammatical and syntactic structures are easier because they are normally use in everyday speech, also in the students' L1. E.g., such phrases as “many”, “more”, “all”, etc.

My belief in aras a universal method of teaching and learning is too firm to be shattered by anything, much less by a single event such as, for example, confirmation of the null hypothesis. I view my Capstone as an opportunity for me to learn how to teach A&D and as such as the beginning of a long journey – a journey that is, by definition, all about the process and not the results. The process is the result.

Because of my belief in A&D as a method of teaching and learning everything (in analogy with The Theory of Everything) is strong, it will not be shattered by the limitations of the study and even if the null hypothesis is confirmed.

Because of the above considerations, my Capstone is process-oriented. The result will be either the rejection of the null hypothesis or its confirmation. Of course, rejection is more desirable. However, even if the null hypothesis is confirmed, it can be reasonably explained by many factors. Limitations, for example: the time constraints, my own lack of experience, the arraignment that I teach not directly but by the proxy of the permanent teacher. These and other possible limitations are enough not to attribute the confirmation of the null hypothesis to the method. On the contrary, it will rather mean that I have to pursue the method further.

So the Capstone is the beginning of that long journey.

February 13, 2018

Debate topic: “Study with or without homework?”

The topic was student-suggested, and the format/setting was also to their most liking: Two teams of four students (flip of a coin) and the teacher acted as a Judge.

80 % of the work with the treatment group will be providing them with templates and structure and the rules – both language- and argument-related:

Talk to the Teacher:

Yes, they know things technical, but the effect of the treatment to a great degree may be attributed to simply having spent much more time and effort during these 10 weeks.

Homework research prior to the discussion is absolutely essential for the academic debate format.

February 15, 2018

Conclusion markers: so, therefore

February 20, 2018

Study with books or digital technology? Student-suggested

Two teams of four (students decide)

No judges

Constructive speeches – ok structure

Books	Digital devices
Intro	Intro
Reason 1: wrong (mistakes)? Reason 2: vision Reason 3: embarrassing if you don't know the authors, especially famous, e.g., H. Tumanyan Additional argument (in discussion/Q&A part: digital devices are expensive, not everybody can afford them Shows lack of prior research/homework (general) Concluding remarks missing	Reason 1: compact Reason 2: unlimited info Reason 3: more interesting No supporting details Additional argument: instant access to vocabulary, no need to use dictionary during reading/studying. Books are past age, digital is new. Concluding remarks - ok

Discussions turn into dialogues: good

As a rule, the students don't do the research part.

February 22, 2018

Paraphrase exercise #2:

the 3rd conditional – into reason and conclusion markers

In doing this exercise, they concentrated on contrasting the concepts, used mostly one linking word of contrast (but) only one instance of using although. Make the exercise easier by showing the contrast linking words.

This paraphrase exercise, together with the paraphrase exercise #1 (class of January 30) were the attempts to use the grammatical structures that were already familiar to the students to teach the important argument markers.

February 27, 2018

Using cars vs. bicycles? Student-suggested

Two teams: Students vs. Teachers

No judges

Tape recorded this time. The students did not seem to mind this time.

This discussion was unique in the sense that the students apparently took it very seriously from the beginning. There were clear signs of homework research done by the students before the debate.

Someone said: So what, your conclusion is missing – about a discussion that sparked about a topic in class. Students start paying attention to the structure of their speech in general – not necessarily during the formal debates.

Cars	Bicycles
<p>1. People absolutely need cars to get around</p> <p>Fast</p> <p>Listen to the music</p> <p>Comfort: door-to-door</p> <p>Exhaust gases and CO2 – exaggeration</p> <p>As the technology grows, A widely adopted route to reduce NO_x emissions is Exhaust Gas Recirculation (EGR). This involves recirculating a controllable proportion of the engine's exhaust back into the intake air.</p>	<p>1. Bicycles are safer. Although many say that bicycles are more dangerous, the numbers do not support this. In US alone, 34,000 people are killed in car accident every year. = a Boeing-747 crashing every week. Will you consider airplanes safe if they crashed every week? Besides, because the cyclists are not protected by the iron box, they are more cautious than car drivers.</p> <p>2. Bicycles are environmentally friendly. (miss intentionally)</p> <p>3. Countries:</p> <p>Bicycles: http://top10hell.com/top-10-countries-with-most-bicycles-per-capita/</p> <p>Happiest: https://www.nationalgeographic.com/travel/top-10/2017-worlds-happiest-countries/</p> <p>Wealthiest: https://www.worldatlas.com/articles/the-richest-countries-in-the-world.html</p> <p>4. Health benefits</p> <p>https://www.betterhealth.vic.gov.au/health/healthyliving/cycling-health-benefits</p> <p>General: It only takes two to four hours a week to achieve a general improvement to your health</p>

	<p>Specific health issues: Obesity and weight control, Cardiovascular disease and cycling, Cancer and cycling, Diabetes and cycling, Bone injuries, arthritis and cycling, Mental illness and cycling</p>
<p>Although ... health issues, but it depends on a person. Many both drive and cycle....</p>	<p>Roads are not made for bicycles. Its us who make the roads</p> <p>Drivers pay for the roads by gas tax (CO2 taxes) so they should get priority. Drivers' payments are not enough, Roads cost 10-15 times more - conservative figure</p> <p>Bicycles are there to stay. Although car-supporters say that cycling is a craze/fad, a short-lived trend of just a fashion, however, cyclists are here to stay, their numbers increase.</p> <p>People absolutely need cars to get around. Perhaps, in low density areas cars are convenient. But in big cities more options are needed. In fact, many people in big cities are both car and bicycle drivers.</p>

	Bicycles per capita % of population drives bicycles	Happiest countries	GDP (PPP) per capita per year (\$) Among the first 30
1	Netherlands -99.1	Netherlands	53,581
2	Denmark -80.1	Denmark	49,613
4	Sweden- 63.7	Sweden (the happiest)	51,264
5	Norway -60.7	Norway	70,590
6	Finland -60.4	Finland	44,050
7	Japan -57		42,659
8	Switzerland -49	Switzerland	61,359
9	Belgium -48		46,301
		Australia	
		New Zealand	
		Canada	
		Iceland	

Students are using and thinking of debate and discussion interchangeably

March 1, 2018

Pros and cons of the debate series: Students' feedback

Free discussion: two teams of students

This discussion was recorded, after the students gave their informed consent. They did not seem to be opposed to the idea of recording anymore. The students' pro and con opinions about the debates are summarized in the table below.

Pro	Con
<ul style="list-style-type: none"> * Vocabulary related to the topics * Vocabulary related to the debate specific terminology * The debate format itself: was new and interesting, especially when the students watched real debates on the internet * The debate where the Harvard debate team lost to the NY prisoners' team 	<ul style="list-style-type: none"> * Appointment of judges from among the students * The bias on the part of the judges * Lack of attention to the current speaker, both from their own team and from the opposing team * The fighting, sometimes even hostile attitude toward the opposing team

Limitations and lessons learned:

- Do not call/label it “debate
- Because the culture of academic debate is non-existent in the Armenian reality, the word debate is what comes to their mind, by association, and they start “debating” i.e., quarrelling/fighting, in a traditional, i.e., hostile way.
- Better to devote time to explain the format, no matter how you call it, the reasons of doing it, e.g., the beauty of a well-organized speech/piece of writing, and then set a debate, simply as a convenient format for discussing controversial issues.

Feedback shared with the students by e-mail after the debate - Students vs. Teachers

Dear Com11 Group -

Thank you for challenging your teachers into the debate. Here are some afterthoughts to share with you.

1. We noticed that you took this debate seriously and were well-prepared for today's discussion by doing your research before the debate.

By the way, to see how a good preparation is important, below are the links to show a debate where the **New York prison inmates** beat the **Harvard University** team.

https://debate.bard.edu/?page_id=2221

<https://www.npr.org/2016/12/27/506314053/after-half-a-century-inmates-resurrect-the-norfolk-prison-debating-society>

<https://www.theguardian.com/education/2015/oct/07/harvards-prestigious-debate-team-loses-to-new-york-prison-inmates>

2. You listened carefully to the constructive speech of the “Cars” team and took advantage of your position of having the 2nd constructive speech. By careful listening you were able to detect the argument of your opponents – the “choice” argument - that was relevant for both sides of the discussion.

3. You probably noticed that “the car” team anticipated a possible argument of the “Bicycle” team – the environmental argument - and how it was included in the “Car” constructive speech. This discussion strategy has a name – it is called “discounting” and shows that you considered many points of view on the same issue.

4. You defended your arguments in the question-and-answer part, for example, bicycles being healthy and less expensive. Your counter-arguments to the “exhaust gases recycling” argument was also very effective.

5. The question-and-answer parts also went well and showed mutual respect and good exchange of information. A kind and respectful exchange of opinions is the best part of any debate.

I hope you enjoyed today’s discussion. Please, think about what you liked and what could be done better next time.

For the next class (the day after tomorrow), please think of some topic or topics to discuss – without preparation, just on the spot.