

EFFECT OF LEARNER PROFICIENCY LEVELS ON METHODOLOGICAL EFFECTIVENESS: CASE OF STAD AND WEBQUEST (STADIBTM)

by **Ghada Awada**

Lebanese American University, P.O. Box 13-5053. Chouran Beirut: 1102 2801. Lebanon

ghadawada @ gmail.com

and **Jack Burston**

Cyprus University of Technology, Limassol 1678, Cyprus

jack.burston @ cut.ac.cy

Abstract

This study set to investigate the critical importance of integrating Student Team Achievement Division (STAD) and *WebQuest*, STADIBTM for short, on developing the advanced-level argumentative writing skills of L2 English university students. The study employed a pre-post-test comparison of the experimental group (N=54) versus the control group (N=24). The data from the study have been systematically reanalysed to evaluate in detail how the initial learner proficiency levels correlated with improvements in the specific areas covered by the evaluation rubric that guided the collaborative writing activities of experimental group students.

Keywords: Proficiency; second language acquisition; Student Team Achievement Division (STAD); *WebQuest*

1. Introduction

It is a standard practice in Second Language Acquisition (SLA) research to evaluate the effectiveness of methodological treatments on the basis of groups of learners, initially determined to be of equivalent proficiency level; the study participants were distributed into the experimental cohort that received the treatment and the control group that did not. Therefore, it is standard practice to measure learning outcomes after a reasonable period based on statistically significant differences in the mean scores of experimental versus control group results on a pre-test/post-test comparison. However, as was shown in Awada, Burston & Ghannage (2019), unless it considers the effect of initial learner proficiency levels, such a procedure can in fact misrepresent actual outcomes. In that study, a comparison of pre-post-test mean score results indicated that the experimental group made statistically significant progress whereas the control group results remained unchanged. Notwithstanding, when the outcomes

were reanalysed based on the division of students into initial low/middle/high proficiency levels, the end results were quite different. It turned out that the positive outcomes were restricted to students with the lowest initial proficiency levels, which was within the control as well as the experimental groups. These results highlighted the importance of going beyond average group results when evaluating the effectiveness of pedagogical methodologies and the need to take into consideration the influence of a range of learner proficiency levels, a largely neglected factor in SLA research design.

It is, thus, the intent of this follow-up study to analyse in greater depth how student interactions in collaborative writing activities in the original study correlated with initial low/middle/high proficiency levels and post-treatment outcomes. This will be done in two stages. Firstly, the structure and results of the original study will be summarized, and then the new data will be presented and analysed.

This study falls within the framework of the constructivism theory (Galloway, 2001) since scaffolding is at the heart of STADIBTM whereby the more skilled learner teaches the less skilled one. This study also falls within the framework of the cooperative learning theory (Johnson, 2013), as the Student Team Achievement Division (STAD) (Slavin, 1990), a cooperative learning structure, is one major component of STADIBTM.

The first backbone of STADIBTM is STAD, a structured cooperative learning method, which demands having one team working together and consisting of heterogenous members in terms of proficiency, activities, and objectives. STAD involves administering individual tests and comparing them to those of the cumulative team scores (Tarim & Akdeniz, 2008; Jolliffe, 2007). STAD demands having a minimum of four and a maximum of five students. This framework encourages individual accountability (Slavin, 1990), promotes learners' understanding and fosters the retrieval of information (Tarim, & Akdeniz, 2008). STAD has been utilized in improving learners' proficiency in mathematics (Rattatumma, & Puncreobutr, 2016) and learners' grammar proficiency as well (Khan & Akhtar, 2017).

WebQuest (Dodge, 1998; Dodge, 2001) is the second component of STADIBTM. WebQuest is inquiry-based learning (IBL) model (Aditomo, Goodyear, Bliuc, & Ellis 2013; Ebadi & Rahimi, 2018; Arsanjani & Faghih, 2015). WebQuest-based instruction involves problem-solving tasks which facilitate the integration of the WebQuest into the classroom and allows learners to effectively employ the gathered information. It facilitates learning from Internet sources that yield in foreign language improvement and cultural exchange in a secure environment (Arsanjani & Faghih, 2015; Awada & Ghaith, 2015; Sysoyev & Evstigneev, 2014). Sox and Rubinstein-Ávila (2009) have asserted the importance of WebQuest for

utilizing authentic and real-life situations. At the same time, Awada and Diab (2018) have reported that WebQuest could improve debate instruction and increase intrinsic motivation towards conducting culturally based debate.

2. Literature review

The theoretical framework of this study was composed of the following five theories. First, Vygotsky's theory of constructivism could explain how learning might occur while learners were actively engaged in meaningful construction of knowledge which is different from passive learning (Galloway, 2001). Second, Student Team Achievement Division (STAD), a cooperative learning structure promotes cooperative learning, improves learners' performance (Slavin, 1990) and enhances learner-centeredness (Johnson, 2013). Third, the learners of this study utilized the WebQuest which could facilitate the project-based learning (PBL) and the inquiry-based learning (Arsanjani & Faghih, 2015). The combination of STAD, a cooperative learning structure, and WebQuest, the Inquiry-Based Technological Model (IBTM), created the framework examined in this study referred to as STADIBTM.

2.1. STADIBTM

This study falls within the framework of the constructivism theory (Galloway, 2001) since scaffolding is at the heart of STADIBTM whereby the more skilled learner teaches the less skilled one. This study also falls within the framework of the cooperative learning theory (Johnson, 2013), as the Student Team Achievement Division (STAD) (Slavin, 1990), a cooperative learning structure, is one major component of STADIBTM.

The first backbone of STADIBTM is STAD, a structured cooperative learning method, which demands having one team working together and consisting of heterogenous members in terms of proficiency, activities, and objectives. STAD involves administering individual tests and comparing them to those of the cumulative team scores (Tarim & Akdeniz, 2008; Jolliffe, 2007). STAD demands having a minimum of four and a maximum of five students. This framework encourages individual accountability (Slavin, 1990), promotes learners' understanding and fosters the retrieval of information (Tarim, & Akdeniz, 2008). STAD has been utilized in improving learners' proficiency in mathematics (Rattanatumma, & Puncreobutr, 2016) and learners' grammar proficiency as well (Khan & Akhtar, 2017).

WebQuest (Dodge, 1998; Dodge, 2001) is the second component of STADIBTM. WebQuest is inquiry-based learning (IBL) model (Aditomo, Goodyear, Bliuc, & Ellis 2013; Ebadi & Rahimi, 2018; Arsanjani & Faghih, 2015). WebQuest-based instruction involves

problem-solving tasks which facilitate the integration of the WebQuest into the classroom and allows learners to effectively employ the gathered information. It facilitates learning from Internet sources that yield in foreign language improvement and cultural exchange in a secure environment (Arsanjani & Faghih, 2015; Awada & Ghaith, 2015; Sysoyev & Evstigneev, 2014). Sox and Rubinstein-Ávila (2009) have asserted the importance of WebQuest for utilizing authentic and real-life situations. At the same time, Awada and Diab (2018) have reported that WebQuest could improve debate instruction and increase intrinsic motivation towards conducting culturally based debate.

2.2. WebQuest inquiry-based instruction

The WebQuest “is not only effective in teaching writing, but it is also effective to improve students’ ability in terms of other language skills” (Adanan, Adanan, & Herawan, 2020, p.78). WebQuest, a highly effective inquiry-based learning (IBL), has facilitated information processing. WebQuest facilitates students’ learning using Internet resources, promotes foreign language skills, and improves cultural awareness in anxiety-reduced environment (Aditomo, Goodyear, Bliuc, & Ellis, 2013; Ebadi & Rahimi, 2018). Furthermore, the WebQuest tasks simulated real-world situations in which each student took on a role related to intercultural communication (Awada & Ghaith, 2015; Sysoyev & Evstigneev, 2014).

In Ebadi & Rahimi’s study (2018) WebQuest-based instruction has improved the critical thinking and academic writing skills of EFL learners enrolled in an IELTS course at EFL institute in Iran. Furthermore, the reflection logs elicited from the learners have shown the learners’ positive experiences towards the effect of the instruction based on the WebQuest (Ebadi & Rahimi, 2018). WebQuest, one of the collaborative activities, has proved to be effective in promoting intercultural communication skills and in increasing motivation when learners are ensured affective, psychological, and cognitive environment (Sigmar, Hynes & Hill, 2012). The WebQuest model showed effectiveness in improving the debate skills, as well as the argumentative and critique writing skills of university learners (Awada and Diab, 2018).

2.3. Effectiveness of STAD Cooperative Learning Structure and STADIBTM

A study conducted by Jahanbakhsh, AliAsgariZamani and Garman (2019) proved the effectiveness of STAD in improving the language proficiency of learners and indicated that STAD is “one of the most popular methods of CL, and has been used to teach different courses ...and has the potential to be used for different purposes” (p. 14). STAD proved to be effective in improving the individual and team tests scores (Tarim & Akdeniz, 2008; Jolliffe, 2007).

Furthermore, STAD has proved its effectiveness in promoting the grammar proficiency of learners (Khan & Akhtar,2017).

STADIBTM consists of STAD, cooperative learning structure, and WebQuest, an inquiry-based learning. The comparison of pre-/post-test mean score results of the experimental group that utilized STADIBTM and the control that employed the regular writing instruction indicated the effectiveness of STADIBTM in making statistically significant progress for the experimental group. Student interactions in collaborative and inquiry-based writing activities combining STAD and WebQuest could affect the low/middle/high proficiency levels of university writing learners (Awada, Burston & Ghannage, 2019). Furthermore, another study conducted by Trianasari and Yuniwati (2019) indicated that “STAD type in the learning process can improve student participation and learning achievement in English language course in the mechanical engineering department” (p. 91). The findings of several studies have shown that “many experts believe that WebQuest can be very useful to help the students in improving their knowledge in various areas of English language skills” (Adanan, Adanan, & Herawan, 2020, p.78).

As such, this study is one of the very few that have intended to contribute to the existing body of knowledge and to the plethora of literature pertaining to the interconnectedness among inquiry-based learning model (IBL), information and communication technology (ICT), cooperative learning approach (CL) and improvement of writing proficiency. In an attempt to bridge the existing gap in literature, this study has been the first to propose STADIBTM. This study aims to analyse in great depth how student interactions in collaborative writing activities could correlate with initial low/middle/high learners’ proficiency levels and post-treatment outcomes.

3. Methodology

3.1. The aim of the study

This study aims at answering these research questions:

1. To what extent did participants in the experimental group discuss the following features covered in the STADIBTM evaluation rubric?
 - a) Creativity
 - b) Content/research
 - c) Citations
 - d) Main points

- e) Organization
- f) Mechanics
- 2. What is the effect of the initial low/middle/high English L2 language proficiency of the experimental group participants upon their advanced-level writing performance as it relates to the above rubric categories?
- 3. How does the writing performance of the initial low/middle/high proficiency level of control students compare to that of the experimental group regarding the above rubric categories?

To achieve the goal, the current study employed the pre-test and post-test experimental design in order to investigate the critical importance of going beyond group averages and taking into consideration a representative range of initial L2 proficiency levels. This investigation took place while evaluating the effectiveness of a pedagogical treatment, a largely neglected factor in Second Language Acquisition research.

3.2. Participants and the research context

This study utilized a convenience sample of 78 English as a foreign language (EFL) participants from six intact sections of EFL university students enrolled in advanced writing course. The native language of the participants was Arabic, and they were of different majors. The STADIBTM experimental group encompassed four sections including 54 participants who were randomly assigned, whereas the control group encompassed two sections including 24 participants. The age of the participants ranged from 18 to 23; 46% of the participants were male learners and 54% were female ones. The average TOEFLiBT entrance score of the experimental group was 81.43, while it was 81.52 for the control group. As such, there was no statistically significant difference between the experimental and the control TOEFLiBT entrance scores. A TOEFLiBT score of 81.5 corresponds to B2+ on the Common European Framework of Reference for Languages scale (Papageorgiou et al., 2015). That is, the score of 81.5 is equivalent to advanced-mid+ on the American Council on the Teaching of Foreign Languages scale (ACTFL.ORG, 2015). The actual range of TOEFLiBT scores for the participants was between 80-86. The pre-post control experimental design was used to collect the quantitative data. Descriptive statistics including the means and standard deviations were utilized to address the study questions.

3.3. Design and procedure

The STAD teams were formed in accordance with the learners' prior knowledge, pre-test achievement scores and the teachers' overall evaluation. As such, the STAD groups were heterogeneous consisting of low, middle, and high achievers. STAD was applied in three steps: First, the experimental teachers delivered whole class instruction; second, the participants received two worksheets to answer in their respective teams in a way that each pair or triad in every team works on every single worksheet. The pairs in every team exchanged worksheets to correct the work of each other and the teachers gave the answer key to all teams to check answers. Third, a test was administered to report achievement (Tarim, & Akdeniz, 2008).

In alignment with the above three steps, the participants in the experimental group sat for a pre-test consisting of an argumentative essay on the effectiveness of intercultural communication. Participants' essays were between 1,250-1,800 words and were written in a three-hour session during which access to supplementary materials including dictionaries and notes was not allowed. The scores of the pre-test were used to identify the three categories: low, middle, and high; the students whose pre-test scores were between 45-68% were identified as low, 69-77% as middle, and over 77% as high achievers. To equalize the sizes of the members' abilities, the cut-off points for the STAD membership were mediated using the teachers' judgement. The participants were then assigned to heterogeneous teams including four participants in twelve groups and three participants in the remaining two.

The whole class approach was utilized to give the instruction for four weeks. The given instruction consisted of lectures on argumentative writing given at a rate of 150 minutes per week. The participants in both groups carried out the tasks using *Writing and Reading across the Curriculum* (12th ed.) for the remaining eight weeks of the semester. Proficient C2 was used as the proficiency required level to effectively carry out the assigned tasks (Council of Europe, 2001, p 27). The Superior level was set for the scores higher than that of Proficient C2 (ACTFL, 2017, p 4). Participants were asked to identify the logical fallacies in a selection and to analyse and synthesize refutations, evidence, means of credibility and rebuttals. Furthermore, they were engaged in debates prior to the argumentative writing exercises and the post-test. All the participants in both experimental and control groups were given two worksheets to answer before the teachers gave them an answer sheet to correct their worksheets for themselves. The worksheets involved drills related to the essay's purpose, organization, coherence, means of credibility, thesis statement, planning, transitional signals, and topic sentences (see Appendix A). The post-test was argumentative essay on the effectiveness of Human Rights Education,

which was between 1,200 and 1,800 words and undertaken in a 150-minute session under the same conditions of the pre-test.

WebQuest was the second component of the treatment. It was created to utilize students' time efficiently and to improve students' analysis, synthesis, problem-solving and evaluation skills (Dodge, 2001). The WebQuest activities required learners to utilize web resources, which were selected ahead of time by the teachers. WebQuests encouraged scaffolding during the learning process since each activity included six sections that helped organize learners' investigations under six main categories: introduction intended to attract the attention and to provide catchy statements and background knowledge, task intended to give the required assignment, resources intended to provide learners with whatever PowerPoint presentations, videos, links or websites, process intended to give the steps to fulfil a task (Iskeceli-Tunc & Oner, 2016), evaluation intended to present the rubric, or the checklist needed to evaluate the task, and conclusion intended to allow learners to reflection on the fulfilled task (Macgregor & McGill, 2005).

Accordingly, the WebQuest created to serve the purpose of this study included an introduction section consisting of articles on Human Rights Education (HRE) integration into school subjects. The task section required students to analyse the HRE articles and to focus on the analysis of claims, counterclaims, support, evidence, means of credibility and rebuttals along with a call to take measures. The process section included the steps needed to carry out the analysis. More importantly, students were asked to freely share their own made videos and the resources they found pertinent to the discussions, debates, and argumentative writing process. Students were also encouraged to post recommendations on the resources they use from WebQuest. The participants were also asked to use the posted rubrics to assess the resources and their writings. The rubric criteria were focused on assessing the content, organization, structure, citations, and creativity.

The instruction stage lasted for the first weeks and was the same for both experimental and control groups whereas the instruction given in the remaining eight weeks differed in accordance with the respective group; the experimental group received the content using STADIBTM while the control group received the regular writing instruction whereby students worked mainly individually and not in groups and did not employ the WebQuest. As such, the STADIBTM instruction was given to the experimental group participants who engaged in their STAD teams to write their thesis statements including claims on HRE integration into school subjects, to use means of support, and make citations. All the experimental participants were asked to outline their essays and draft their essay body, introduction, and conclusion in groups.

Furthermore, students were engaged as well in revising and editing their essays. The STAD members exchanged and discussed the worksheet answers and eventually they compared their answers to the answer keys given by the teachers.

In contrast, following the initial course lectures, the control group participants continued to receive argumentative writing instruction for eight weeks through lecturing and whole class teacher-student discussion. The instruction given to the control group included guidelines on how to develop affirmative and negative positions, how to draft an outline of argumentative writing, provide affirmative statements, draft cross examination, and locate and identify sources needed for their writings. The control group participants completed the same worksheets as the experimental group. However, unlike the STADIBTM students, the control group students were responsible for finding their own online complementary materials. Whereas the experimental group students always worked collaboratively, most control group students worked individually even though they were free to work in pairs or small groups. Unlike the experimental group which received peer and instructor feedback, the control group received only instructor feedback.

3.4. Data collection tools and procedures

A pre-test (see Appendix C) and a post-test (see Appendix D) along with a rubric were the three instruments used to collect data to address the questions raised by this study. The control and the experimental groups undertook the pre- and post-test essays that were scored numerically by utilizing one rubric (see Appendix B). Reliability of correction was maintained by having all essays corrected by two teachers. A third teacher intervened in the case of a significant discrepancy (+/- 5-7 points) to attain a consensus. The marking was completely blind to avoid any bias towards any of the two groups. ANCOVA statistics were carried out to obtain quantitative data including the means and standard deviations of the scores of the pre and post-tests.

In the original study, a quantitative ANCOVA data analysis was conducted in two stages. In the first step, the means and standard deviations of the whole experimental group were compared with those of the whole control on the pre-treatment and post-treatment essays. The purpose of this analysis was twofold. Firstly, it provided a baseline against which progress could be measured and secondly it played a major role in demonstrating the equivalence of L2 English language proficiency between the experimental and control groups.

In the second stage, the analysis compared pre-/post-treatment scores within the experimental and control groups relative to the low/middle/high proficiency groupings of the

participants. The purpose of this analysis was to determine the extent to which the effectiveness of the STADIBTM methodology was affected by the initial L2 English proficiency level of students.

The quantitative ANCOVA results of the pre-treatment essay for the experimental and control groups showed a higher mean for the latter (68.33333 –SD 11.23143) compared to the former (64.83333 –SD 11.42449), however, this difference was not statistically significant ($p > 0.05$), as shown in Table 1. This, in conjunction with the students' TEFLiBT scores, confirmed the initial equivalence of the experimental and control groups.

Table 1. Paired sample test

		Mean	Std. Deviation	Std. Error Mean	Sig. (2-tailed)
Pair 1	Pretest cont_L – Pretest exp_L	1.33333	13.95230	5.69600	.824
Pair 2	Pretest cont_M Pretest exp_M	1.53846	2.96129	.82131	.086
Pair 3	Pretest cont_H Pretest exp_H	-.60000-	4.97996	2.22711	.801

Table 2 shows the mean scores, standard deviations, and the statistical significance ($p < 0.05$) of the difference between the mean scores of the pre-/post-test between the experimental and control groups. As can be seen, whereas the global post-test results of the experimental group significantly improved by almost 10 points (64.8333 –SD 11.42449 / 74.5185 –SD 11.88743), those of the control group remained essentially unchanged (68.3333 –SD 11.23143 / 68.9583 –SD 13.26807).

Table 2. Paired samples statistics

Pair 1	pretestexp_total	64.8333	54	11.42449	1.55468	.000
	posttestexp_total	74.5185	54	11.88743	1.61767	
Pair 2	pretestcont_total	68.3333	24	11.23143	2.29261	.859
	posttestcont_total	68.9583	24	13.26807	2.70833	

Table 3 shows the mean scores, standard deviation and the statistical significance ($p < 0.05$) of the difference between the mean scores for each of the low/middle/high groups on the pre-/post-treatment essays for the experimental and control groups.

Table 3. Paired samples statistics

		Mean	N	Std. Deviation	Std. Error Mean	p Value
Pair 1	pretestexp_L	53.5263	19	10.02424	2.29972	.000
	posttestexp_L	72.5789	19	14.45420	3.31602	
Pair 2	pretestexpM	70.4545	22	.91168	.19437	.230
	posttestexp_M	73.5909	22	11.82676	2.52147	
Pair 3	pretestexp_H	79.7143	7	4.57217	1.72812	.267
	posttestexp_H	81.7143	7	3.68394	1.39240	
Pair 4	pretestcont_L	51.8333	6	9.76559	3.98678	.002
	posttestcont_L	71.0000	6	6.13188	2.50333	
Pair 5	pretestcont_M	72.0000	13	2.61406	.72501	.250
	posttestcont_M	66.2308	13	17.22960	4.77863	
Pair 6	pretestcont_H	78.6000	5	1.34164	.60000	.071
	posttestcont_H	73.6000	5	4.15933	1.86011	

Among the experimental group participants, the post-test scores of the lower ability experimental group students (N=19) increased significantly by over 19 points (SD 14.45420). On the other hand, there were no significant gains among the middle (N=22) or high (N=7) ability students. In fact, what little difference there was correlated inversely with ability level, +3.1364 (SD 11.82676) for middle ability and + 2.0000 (SD 3.68394) for high ability. Likewise, within the control group, while the scores of middle ability students (N=13) decreased by 5.7692 (SD 17.22960) and those with higher ability (N=5) increased by +5.0000 (SD 4.15933), neither of these changes were statistically significant. On the other hand, for low ability students (N=6) there was a significant difference of over 19 points (SD 4.15933) between pre-/post-test results. Lastly, comparing the pre-/post-test results of the lower ability students in the experimental and control groups (See Table 4), there was no significant difference. Neither was there any significant difference between the middle ability learners in the experimental and control groups nor between the high ability learners in the two groups.

Table 4. Paired samples statistics

		Mean	N	Std. Deviation	Std. Error Mean	p Value
Pair 1	exp_diff_L	28.1667	6	15.71517	6.41569	.289
	cont_diff_L	19.1667	6	7.78246	3.17718	
Pair 2	exp_diff_M	10.2308	13	7.28187	2.01963	.862
	cont_diff_M	11.0000	13	14.17745	3.93212	
Pair 3	exp_diff_H	3.6000	5	4.33590	1.93907	.435
	cont_diff_H	5.8000	5	3.19374	1.42829	

As was demonstrated in the original STADIBTM study, when evaluating the effectiveness of a pedagogical treatment based on pre-test/post-test results, it is critically important to go beyond group averages and take into consideration a representative range of initial L2 proficiency levels. Because of the way STADIBTM activities were organized, it is in fact possible to obtain a much fine-grained analysis of learning outcomes relative to student proficiency levels. In the STADIBTM study, all participants were made aware, and had a copy of, the evaluation rubric used in the assessment of their essays. As part of the WebQuest procedures, the students in the experimental group applied the rubric not only to their evaluation of the resources they read but also to the preliminary drafts they themselves collaboratively wrote. Unlike the experimental group participants, it was left for those in the control group to use the rubric as they saw fit. While in the original STADIBTM study the rubric formed the basis of the evaluation of the pre-/post treatment essays, it was used only to provide a global assessment of writing performance in terms of the aggregate scores that derived from its various categories. However, since the STADIBTM participants systematically used the rubric to guide their writing, the effect of this approach can be more informatively assessed category by category in the pre-/post-treatment essays, and this relative to the initial low/middle/high proficiency level of the experimental and control group learners.

4. Findings and discussion

4.1. Findings on research questions 1 and 2

Addressing questions 1 and 2 required summarizing by low/middle/high proficiency level the interactive discussions of students concerning the aspects of the rubric. When dealing with the question of discussion topics, it was necessary to go through the transcripts of the student

interactions in order to determine firstly the overall percentage that deals explicitly with the evaluation metrics, then to see if there was any correlation between the three student proficiency levels and participation in those rubric-related discussions. Determining the effect of initial proficiency level upon learning outcomes required an analysis of the pre-treatment and post-treatment essays of the six rubric categories for each of the low/middle/high proficiency levels. This needed to be done internally to the experimental and control groups. It was also necessary to compare the results of the experimental and control groups, low/middle/high level by level.

1. Extent to which participants in the experimental group discussed the STADIBTM evaluation rubric

Analysis of the student discussions indicated that the middle-level students were raising more questions than the lower-level students and that the higher-level students were answering them. That is, most of the discussion was between middle and high-level students, yet discussion among low, middle and high-level students did also take place. Secondly, the analysis showed that a breakdown of the discussions by specific rubric category was possible. Specifically, 60% of the student communications referred to the rubric, of which 25% dealt with content, 15% creativity, 15% mechanics, 15% main points, 10% citations, 20% organization. Furthermore, the findings showed that there was a pattern to the interactions related to proficiency level (see Table 5). The low-level students concentrated on content, mechanics and main points, middle-level ones on creativity and content, high-level students on organization and citations. Thirdly, the findings indicated that the collaborative writing that was produced aligned to a significant extent with the discussion of the rubric categories which was reflected in actual changes made to drafts. Thus, 75% of changes relating to creativity and content were made by middle-level students and 25% of the content changes were made by low-level students. 55%, 40% and 5% of creativity changes were made by high, middle, and low-level students, respectively. The significant citation and organization changes were made by high level students.

Table 5. Experimental and control group post-test results

Proficiency level	High		Middle		Low	
	Con Post-test	Exp Post-test	Con Post-test	Exp Post-test	Con Post-test	Exp Post-test
Rubric Aspects						

Creativity	45%	40%	35%	42%	20%	18%
Content/research	40%	35%	40%	40%	20%	25%
Citations	40%	55%	35%	30%	25%	10%
Main points	40%	45%	35%	35%	25%	20%
Organization	40%	50%	35%	35%	25%	15%
Mechanics	45%	45%	35%	30%	20%	25,00%

2. Experimental group pre-/post-test results

For the second research question, the results of the pre-treatment and post-treatment experimental group essays were compared by rubric category for each of the three proficiency levels. As can be seen in Table 6 the level of improvement was inversely proportional to initial proficiency level for all rubric categories. The lowest level students improved by 5% to 10% in all categories. Except for a slight 2% increase in creativity, middle-level students either made no progress or retrograded by 5%. High-level students improved the least, making no progress in citations and retrograding by 5% -10% in everything else. It must be kept in mind, however, that the results of both the middle and high-level students were higher from the beginning and remained so after the treatment.

Table 6. Experimental group pre-/post-test results

Proficiency level	High		Middle			Low			
	Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test	
Rubric Aspects									
Creativity	50%	40%	-10%	40%	42%	+2%	10%	18%	+8%
Content/research	45%	35%	-10%	40%	40%	0%	15%	25%	+10%
Citations	55%	55%	0%	35%	30%	-5%	5%	10%	+5%
Main points	55%	45%	-10%	35%	35%	0%	10%	20%	+10%
Organization	55%	50%	-5%	35%	35%	0%	10%	15%	+5%
Mechanics	50%	45%	-5%	35%	30%	-5%	15%	25%	10.00%

3. Control group pre-/post-test results

For the second research question, the researchers needed to compare the results of the control pre-treatment and post-treatment essays by rubric category for each of the three proficiency levels. As can be seen in Table 7, except for mechanics which remained unchanged, the control high-level students regressed in all categories. Similarly, the control middle-level students made no progress in half of the categories and retrogressed in the other half. On the other hand, the control low level students showed significant progress of between 5% and 15% across the board. The difficulty level of the topic and the

fact that the grades of the high-level students were higher at the beginning could explain the findings, as shown in Tables 7 & 8

Table 7. Control group pre-/post-test results

Proficiency level	High		Middle			Low			
	Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test	
Rubric Aspects									
Creativity	50%	45%	-5%	40%	35%	-5%	10%	20%	+10%
Content/research	45%	40%	-5%	40%	40%	0%	15%	20%	+5%
Citations	50%	40%	-10%	35%	35%	0%	15%	25%	+10%
Main points	50%	40%	-10%	35%	35%	0%	15%	25%	+10%
Organization	50%	40%	-10%	40%	35%	-5%	10%	25%	+15%
Mechanics	45%	45%	0%	40%	35%	-5%	15%	20%	+5%

4.2. Findings on question 3: How does the writing performance of the initial low/middle/high proficiency level of control students compare to that of the experimental group regarding the above rubric categories?

The comparison between the writing performance of the initial low/middle/high proficiency level of control students and those of the experimental group regarding the above rubric categories does not show significant differences.

40%, 35%, 55%, 45%, 50%, and 45% were achieved in terms of creativity, content, citations, main points, organization, and mechanics, respectively. The control high level students scored 45%, 40%, 40%, 40%, 40% and 45% on creativity, content, citations, main points, organization, and mechanics, respectively.

The experimental middle level students scored 42%, 40%, 30%, 35%, 35%, and 30% on creativity, content, citations, main points, organization, and mechanics, respectively. The control middle level students scored 35%, 40%, 35%, 35%, 35% and 35% on creativity, content, citations, main points, organization, and mechanics, respectively.

The experimental low-level students scored 18%, 25%, 10%, 20%, 15% and 25% on creativity, content, citations, main points, organization, and mechanics, respectively. The control low level students scored 20%, 20%, 25%, 25%, 25% and 20% on creativity, content, citations, main points, organization, and mechanics, respectively.

Table 8. Experimental and control initial low/middle/high proficiency level of experimental and control students

Proficiency level	High			Middle			Low		
	Exp	Con		Exp	Con	%	Exp	Con	%
Rubric Aspects									
Creativity	40%	40%	0%	40%	42%	%	5%	10%	%
Content/research	45%	45%	0%	40%	35%	%	15%	15%	0%
Citations	40%	45%	%	35%	35%	0%	25%	15%	%
Main points	40%	45%	%	35%	35%	0%	25%	15%	%
Organization	55%	50%	%	35%	35%	0%	10%	10%	%
Mechanics	55%	55%	0%	30%	35%	%	15%	15%	%

4.3. Discussion and implications

As was demonstrated in the original STADIBTM study, when evaluating the effectiveness of a pedagogical treatment based on pre-test/post-test results, it is critically important go beyond group averages and take into consideration a representative range of initial L2 proficiency levels. Because of the way STADIBTM activities were organized, it is in fact possible to obtain a much finer-grained analysis of learning outcomes relative to student proficiency levels. In the STADIBTM study, all participants were made aware, and had a copy of, the evaluation rubric used in the assessment of their essays. As part of the WebQuest procedures, students in the experimental group applied the rubric not only in their evaluation of the resources they read but also in relation to the preliminary drafts they themselves collaboratively wrote. Unlike the experimental group participants, it was left for those in the control group to use the rubric as they saw fit. While in the original STADIBTM study the rubric formed the basis of the evaluation of the pre-/post treatment essays, it was used only to provide a global assessment of writing performance in terms of the aggregate scores that derived from its various categories. However, since STADIBTM participants systematically used the rubric to guide their writing, the effect of this approach can be more informatively assessed category by category in the pre-/post-treatment essays, and this relative to the initial low/middle/high proficiency level of the experimental and control group learners.

In the original study, a quantitative ANCOVA data analysis was conducted in two stages. In the first step, the means and standard deviations of the whole experimental group were compared with those of the whole control on the pre-treatment and post-treatment essays. The purpose of this analysis was twofold. Firstly, it provided a baseline against which progress could be measured and secondly it played a major role in demonstrating the equivalence of L2 English language proficiency between the experimental and control groups. In the second stage, the analysis compared pre-/post-treatment scores within the experimental and control groups relative to the low/middle/high proficiency groupings of the participants. The purpose of this analysis was to determine the extent to which the effectiveness of the STADIBTM methodology was affected by the initial L2 English proficiency level of students.

Findings indicated that the middle-level students were raising more questions than the lower-level students and the higher-level students were answering them. That is, most of the discussion was between middle and high-level students, yet there was a significant discussion among low, middle, and high levelled students. In fact, there was a discernible pattern related to proficiency level. Secondly, the study findings showed that a breakdown of the discussions by specific category could be provided. For example, 60% of the student communications referred to the rubric, of which 25% dealt with content, 15% creativity, 15% mechanics, 15% main points, 10% citations, 20% organization. Furthermore, the findings asserted that there was a pattern to the interactions related to proficiency level; for example, the low-level students concentrated on mechanics and main points, middle-level ones on structure and organization and high-level students on content and citations. Thirdly, the findings indicated that the collaborative writing that was produced aligned to a significant extent with the discussion of the rubric categories which was reflected in the actual changes made to drafts. As such, 75% of the changes relating to mechanics were made by mid-level students and 25% of the content changes were made by low-level students, 55%, 40% and 5% of creativity changes were made by high, mid and low-level students respectively. 40%, 35% and 25% of the citation changes were made by mid, high, and levelled students, respectively.

Low-level students improved by 5% on creativity, 10% on content, 10% on citations, 15% on main points, 5% on organization, and 5% on mechanics. As such, the low-level students showed slight improvement in creativity, mechanics, and organization. The middle-level students improved by 5% on creativity, 5% on content, 10% on citations, 10% on main points, 5% on organization, and 10% on mechanics. As such, the middle-level students showed slight improvement in creativity, content, mechanics, and organization. The high-level students retrogressed by 10% on creativity, 15% on content, 15% on citations, 15% on main points, 10% on organization, and 15% on mechanics. As such, the high-level students did not show improvement in comparison with the low and middle-level students. However, the low and middle-level students showed improvement. Yet, the difficulty level of the topic and the fact that the grades of the high-level students were high from the beginning and remained so after the treatment could explain the findings. There were correlations between what was discussed/practiced during the STADIBTM student interactions and the post-test results. In fact, what was discussed and practiced had a flow-on effect in the post-treatment essay.

The experimental group with low-level students showed no improvement on creativity, citations and mechanics. They retrogressed by 5% on content, and retrogressed by 5% on main points, and retrogressed by 5% on organization. The middle-level students showed no

improvement on citations and mechanics. They improved by 5% on creativity, 5% on content, any 5% on organization. However, they retrogressed by 5% on main points. As such, the middle-level students showed improvement in creativity, content, mechanics, and organization. The high-level students retrogressed by 10% on creativity, improved by 5% on content, 15% on citations, 15% on main points, 10% on organization, and 15% on mechanics. As such, unlike the low and middle-level students, the high-level students did not show improvement on creativity. Yet, the difficulty level of the topic again and the fact that the grades of the high-level students were high from the beginning and remained so after the treatment could explain the findings pertinent to the high-level students. Yet, the low-level students of the control group did not show improvement, which indicated a stronger correlation between the effectiveness of the STADIBTM and the improvement the experimental low-level students showed.

The findings of the study aligned with those of Tarim and Akdeniz (2008) and Jolliffe (2007) that indicated the effectiveness of STAD as a cooperative learning method in which each team member cooperates in ensuring that partners have mastered assigned objectives based on individual tests and team scores. The study results corroborate those of Khan and Akhtar (2017) that demonstrated the effectiveness of STAD in improving the grammar and mechanics proficiency of learners. Furthermore, there is an alignment between the findings of the study and those of Aditomo, Goodyear, Bliuc, and Ellis (2013), as well as Ebadi and Rahimi (2018), which showed that WebQuest is a well-established model for inquiry-based learning (IBL) which encouraged students to focus on how to find and employ information. Furthermore, the study findings corroborate those of Awada and Ghaith (2015) and Sysoyev and Evstigneev (2014), which indicated that the WebQuest tasks simulated real-world situations in which each student took on a particular role related to intercultural communication.

In the same vein, the study findings corroborate those of Ware (2013) and Shuter (2012) that indicated that there is a positive interaction between intercultural communication skills and collaborative technology-based activities such as the WebQuest. Furthermore, the findings aligned with those of Sigmar, Hynes and Hill (2012) that asserted that learners could enhance their intercultural communication skills and increase their motivation should they are ensured affective, psychological, and cognitive environment. More importantly, the findings of this study corroborate with those of Awada and Diab (2018) that indicated the significance of the WebQuest media in improving the debate skills of learning and possibly enhancing their argumentative and critique writing skills.

5. Conclusion

This study was a sequel to that presented in Awada, Burston and Ghannage (2019), which investigated the effectiveness of integrating Student Team Achievement Division (STAD) and WebQuest, STADIBTM for short, on developing the advanced-level argumentative writing skills of EFL university students. Findings indicated that most of the discussion was between middle and high-level students, yet there was a significant discussion among low, middle, and high-level students. In fact, there was a discernible pattern related to proficiency level. The middle-level students showed slight improvement in creativity, content, mechanics, and organization. The high-level students did not show improvement in comparison with the low and middle-level students. Secondly, the study findings showed that a breakdown of the discussions by specific category could be provided.

This study also yielded important pedagogical implications. First, it is more vital for students to work in heterogenous groups and not individually to fulfil a writing task in general and argumentative writing requirements. Second, it is imperative that writing teachers utilize peer and teacher feedback to improve low-level students' writing proficiency. There is also a positive interaction between intercultural communication skills and collaborative technologies such as the WebQuest. Future research should be conducted to investigate the impact of using STADIBTM on learners' motivation and to examine whether there is a correlation between using STADIBTM, improving learners' writing proficiency and increasing motivation.

Acknowledgement

I would like to thank the editor, Professor Jarek Krajka, and the two anonymous reviewers, for their constructive comments on my paper.

References

- ACTFL.ORG, (2015). *Assigning CEFR ratings to ACTFL assessments*. Alexandria, VA: ACTFL. Retrieved 19 July, 2020, from https://www.actfl.org/sites/default/files/reports/Assigning_CEFR_Ratings_To_ACTFL_Assessments.pdf
- Adanan, H., Adanan, M., & Herawan, T. (2020). M-WebQuest development: reading comprehension of senior high school students in Indonesia. *International Journal of Emerging Technologies in Learning (iJET)*, 15(3), 74-92.
- Aditomo, A., Goodyear. A., Bliuc, M., & Ellis, R. (2013). Inquiry-based learning in higher education: Principal forms, educational objectives, and disciplinary variations. *Studies in Higher Education*, 38(9), 1239-1258. <http://dx.doi.org/10.1080/03075079.2011.616584>

- Arsanjani, M., & Faghieh, E. (2015). The impact of the *WebQuest* instruction system on Iranian intermediate EFL learners' writing performance and perception. *International Journal of Instructional Technology and Distance Learning*, 12(2), 37-47.
- Awada, G., Burston, J., & Ghannage, R. (2020). Effect of student team achievement division through WebQuest on EFL students' argumentative writing skills and their instructors' perceptions. *Computer Assisted Language Learning*, 33(3), 275-300. DOI: 10.1080/09588221.2018.1558254
- Awada, G., & Diab, H. (2018). Effect of WebQuest media on learners' intercultural communication debate and motivation. 19th International CALL Conference: CALL your Data, Bruges, Belgium. https://www.researchgate.net/publication/343080738_CALL_Paper-Awada_and_Diab2018
- Awada, G. & Ghaith, G.(2015). Impact of WebQuest and gender on writing achievement in professional business English. *Taiwan International ESP Journal*, 6(2), 1-27. DOI: 10.6706/TIESPJ.2014.6.2.1
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press. Retrieved 19 July, 2020, from <https://rm.coe.int/1680459f97>.
- Dodge, B. (2001). FOCUS: Five rules for writing a great WebQuest. *Learning and Leading with Technology*, 28(8), 6-9. Retrieved 19 July, 2020, from http://www.academia.edu/download/50405963/Five_20Rules_20for_20Writing_20a_20Great_20WebQuest.doc
- Dodge, B. (1998). WebQuests: A strategy for scaffolding higher level learning. Paper presented during *Educational Computing Conference*, San Diego, CA. Retrieved 19 July, 2020, from <http://edweb.sdsu.edu/webquest/necc98.htm>
- Ebadi, S., & Rahimi, M. (2018). An exploration into the impact of WebQuest-based classroom on EFL learners' critical thinking and academic writing skills: A mixed-methods study. *Computer Assisted Language Learning*, 31(5-6), 617-651. DOI: 10.1080/09588221.2018.1449757
- Galloway, C. A. (2001). Vygotsky's learning theory. Emerging Perspective on Learning, Teaching and Technology. Retrieved 19 July, 2020, from <http://www.coe.uga.edu/epltt/vygotskyconstructivism.htm>
- Iskeceli-Tunc, S., & Oner, D. (2016). Use of WebQuest design for inservice teacher professional development. *Education and Information Technologies*, 21(2), 319-347. Retrieved 19 July, 2020, from <https://link.springer.com/article/10.1007/s10639-014-9323-y>
- Jahanbakhsh, A. A., AliAsgariZamani, M., & Garman, Z. (2019). CIRC and STAD in Iranian context: Through the five elements to cooperative learning of lexical collocations. *Cogent Arts & Humanities*, 6(1),1-17. <https://doi.org/10.1080/23311983.2019.1692469>
- Johnson, E. (2013). *The Student-Centered Classroom: Vol 1: Social Studies and History*. New York(NY). Routledge.
- Jolliffe, W. (2007). *Cooperative Learning in the Classroom: Putting It into Practice*. London: Sage.
- Khan, A., & Akhtar, M. (2017). Investigating the effectiveness of cooperative learning method on teaching of English grammar. *Bulletin of Education and Research*, 39(1), 1-16. Retrieved 19 July, 2020, from <https://search.proquest.com/docview/1944195134?accountid=27870>

- Macgregor, G., & McGill, L. (2005). Digital libraries and information literacy issues within virtual learning environments: An e-learning impasse? Paper presented at *Librarians' Annual Information Literacy Conference (LILAC)*. Retrieved 19 July, 2020, from <http://strathprints.strath.ac.uk/2334/>
- National Council of State Supervisors for Languages & American Council on the Teaching of Foreign Languages (2017). *NCSSFL-ACTFL Can-Do Statements: Performance Indicators for Language Learners*. Retrieved 19 July, 2020, from https://www.actfl.org/sites/default/files/CanDos/Superior%20Distinguished%20Can-Do_Statements.pdf
- Rattanatumma, T., & Puncreobutr, V. (2016). Assessing the effectiveness of STAD model and Problem Based Learning in mathematics learning achievement and problem-solving ability. *Journal of Education and Practice*, 7(12), 194-199. EJ1099565
- Papageorgiou, S., Tannenbaum, R., Bridgeman, B., & Cho, Y. (2015). The association between TOEFL iBT® test scores and the Common European Framework of Reference (CEFR) levels. (Research Memorandum ETS RM-15-06). Princeton, NJ: Educational Testing Service. Retrieved 19 July, 2020, from <https://www.ets.org/Media/Research/pdf/RM-15-06.pdf>
- Shuter, R. (2012). Intercultural new media studies: The next frontier in intercultural communication. *Journal of Intercultural Communication Research*, 41(3), 219-237.
- Sigmar, L. S., Hynes, G. E., & Hill, K. L. (2012). Strategies for teaching social and emotional intelligence in business communication. *Business Communication Quarterly*, 75(3), 301-317.
- Slavin, R. (1990). *Cooperative Learning: Theory, Research, and Practice*. Englewood Cliffs, NJ: Prentice-Hall.
- Sox, A., & Rubinstein-Ávila, E. (2009). WebQuests for English-language learners: Essential elements for design. *Journal of Adolescent & Adult Literacy*, 53(1), 38-48.
- Sysoyev, P. & Evstigneev, M. (2014). Foreign language teachers' competency and competence in using information and communication technologies. *Procedia-Social and Behavioral Sciences*, 154, 82-86. <https://doi.org/10.1016/j.sbspro.2014.10.116>
- Tarim, K., & Akdeniz, F. (2008). The effects of cooperative learning on Turkish elementary students' mathematics achievement and attitude towards mathematics using TAI and STAD methods. *Educational Studies in Mathematics*, 67(1), 77-91. Retrieved 19 July, 2020, from <https://link.springer.com/article/10.1007/s10649-007-9088-y>
- Trianasari, E., & Yuniwati, I. (2019). Application model of student team achievement division (stad) in English class. *Journal of Applied Studies in Language*, 3(1), 91-99. DOI: <http://dx.doi.org/10.31940/jasl.v3i1.1088>
- Ware, P. (2013). Teaching comments: intercultural communication skills in the digital age. *Intercultural Education*, 24(4), 315-326.

Appendices

Appendix A: Rubric

Scoring Rubric for Short, Research-based Position Paper

Undergraduate Cognitive Psychology, Anne L. Fay, Carnegie Mellon University.

<https://www.cmu.edu/teaching/resources/Teaching/CourseDesign/Assessment-Grading/Rubrics/PsychologyPaperRubric.doc>

Appendix B: Argumentative Writing Material

Adapted from: Sinnott-Armstrong, W., & Fogelin, R. J. (2014). *Cengage Advantage Books: Understanding arguments: An introduction to informal logic*. Cengage Learning.

http://zu.edu.jo/UploadFile/Library/E_Books/Files/LibraryFile_17113_11.pdf

Appendix C: Pretest

Students were asked to argue with or against the integration of Human Rights Education into all school subjects. Students were instructed to provide an introduction, summary of utilized references concessions, refutations, support paragraphs, call to action and a conclusion.

Appendix D: Posttest

Students were asked to argue with or against the integration of Human Rights Education into all university courses. Students were instructed to provide an introduction, summary of utilized references concessions, refutations, support paragraphs, call to action and a conclusion.