TITLE: COOPERATIVE TESTING WITH FLIPPED CLASSROOM METHODS FOR 9TH GRADE BIOLOGY STUDENTS

ABSTRACT

Cooperative groups are known to be effective learning tools for students in the classroom. Over 500 studies were conducted between 1898 and 1989 on cooperative, competitive, and individualistic efforts (Johnson et al., 2000). The cognitive, social, and academic benefits are numerous for working in groups. However, when the time comes for formative assessment, most teachers separate students to take tests individually. The activity described in this article offers an alternative to individual assessment taking while at the same time increases academic achievement and motivation in science courses. In conjunction with the flipped classroom method, this activity engaged students in the learning process. It also helped springboard their affinity for science that can be applied to any topic of study or grade level (with slight modifications for younger grade levels). Three sections of a 9th grade biology class were used for this example which included 66 students for 13 weeks with class periods of 50 minutes.

Keywords: Science, flipped classroom, cooperative learning, assessment

1. INTRODUCTION

The United States ranked 27th and below in secondary science education on the Programme for International Student Assessment (PISA) tests in 2012 and 2006 (OECD, 2012). It is a startling concept to think the U.S. is lagging so far behind so many other countries in the world in science education according to these tests. Many think a technique called flipped instruction is an answer to this problem (Berrett, 2012). Flipped instruction is a method of teaching that has created quite a buzz in the educational world (Brame, 2013). Flipped classrooms involve the use of a short, instructional video to introduce a topic. Students complete the video assignment as homework in order for time in the classroom to be used investigating the topic together in groups. When students return to class with their homework from the video lesson, they typically take a short assessment for the teacher to verify learning. This creates a more student-centered classroom and enables the teacher to engage students with deeper content knowledge instead of using precious class time for lectures as in traditional settings. Many schools have had great success with flipped classrooms and teachers all over the world have instituted this instructional method. However, there does not seem to be sufficient research to support a widespread paradigm shift to...
flipped learning (Kwan & Khe, 2017). Cooperative learning is a key component of the flipped classroom with students working together in class more so than in traditional classrooms. Cooperative learning has been extensively researched which reveals its benefits (Johnson & Johnson, 2013). Therefore, this study combines flipped instruction with cooperative learning and testing and investigated effects on academic achievement, motivation toward science, and study time.

The study investigated the effects of cooperative testing versus individual testing with flipped classroom lessons in 9th grade biology classes on academic achievement, motivation toward science and study time. Flipped learning and cooperative testing as successful instructional strategies have been used separately in numerous studies, articles, and classrooms (Johnson & Johnson, 1999; Hamdan, et al., 2013; Roediger & Karpicke, 2006). However, these strategies have seldom been used together to determine effectiveness nor featured in a high school setting.

During cooperative learning and cooperative testing, teachers move around the room and assist students with any misconceptions they may hear students discussing. When a teacher is merely directing a class, he or she is not getting necessary feedback from the cognitive processes taking place for the student. Cooperative learning enables students to bounce off ideas of others for feedback. Students are also more likely to ask questions in small group settings rather than large, traditional, classroom discussion settings which can be intimidating for many. Because this method has the students working in small groups that become strengthened through time spent in lab activities and collaborative testing, the students are able to share their thoughts more readily. When using cooperative learning educators are able to assist with any hurdles the students may encounter through their discussions and resolve any misconceptions. In science, misconceptions are often unnoticed because students do not have opportunities to describe their understanding of the information or topics as fully as they do in a flipped, cooperative setting.

2. METHODS and MATERIALS

This study was conducted in an on-campus, face-to-face, three sections of 9th grade biology class (38 males and 28 females) in a private high school. This biology course was a required science credit for graduation. Student ability levels varied. Students were provided with instruction on flipped learning and note taking requirements for preparation of the weekly, 10-point quizzes. Direct feedback was provided on quizzes and classroom activity grades throughout the course. None of the students had prior experience with cooperative testing methods, but all had prior experience with cooperative groups for projects and other assignments in other classes.

Two experimental groups received treatment of either a cooperative testing or individual testing quiz each week for 12 weeks (Week 7 was reserved for the mid-study questionnaire administration and no quiz was given). The cooperative testing groups consisted of groups of three or four students who were selected at random at the beginning of the study and remained in
the same group throughout the 13-week study. The method of test administration (cooperative or individual) was also selected at random to provide six cooperative quizzes and six individual quizzes for each of the three classes. Students did not know ahead of time if the quiz would be individual or cooperative.

Dependent variables were the scores on the weekly quizzes for academic achievement, final exams for the semester for student retention of information learned, impacts on motivation in science as measured by the questionnaires, study time on particular topics, and effectiveness of cooperative group exercises. A questionnaire collected the dependent variable information on cooperative learning impacts on motivation toward science. The questionnaire was distributed at the beginning of the study, in week 7 (middle), and end of the study during week 13. Additionally, two questionnaire items were included after their weekly quizzes for data collection of their study time for the week and if students felt this was sufficient study time based on the quiz. Semi-structured interviews with 10 students and one focus group of 10 students were administered in the final week of the research study.

Instructional materials were based on the weekly lessons in the biology course. The flipped learning instruction material consisted of a homework assignment of watching and taking notes on a 5 to 12-minute instructional video of content material. Instructional videos were taken from various Internet-based sources such as Khan Academy, Crash Course, and Amoeba Sisters for biology lessons. The same quiz was administered to each class, but the method of test taking varied from cooperative to individual, as selected randomly for each class for the week. Oftentimes students rely on other group members to do the work for cooperative learning, which results in a concept known as “social loafing” (Smrt& Karau, 2011).

3. RESULTS

Research question 1. Does the use of cooperative testing increase students ‘academic achievement in a 9th grade biology course as measure by quiz scores over the 13-week study?

In order to determine whether these differences were truly from the group effort working together or merely the highest scoring individual in the group, simple t-tests were conducted to compare the highest scores of individual test scores with the highest cooperative scores. Three of the 12 weeks of quizzes administered had the same format (all cooperative or all individual quizzes) for all three classes and were not included in the comparison because there was no variation. Therefore only nine of the 12 quizzes were compared. Week 3 all classes were cooperative, all classes were tested individually in week 4, and all were tested cooperatively in week 5. Table 1 lists the means and standard deviations of quiz scores by treatment group and by best scores of those treatment groups.

Table 1

Means, Standard Deviations, and t-tests of Quiz Scores
<table>
<thead>
<tr>
<th>Quiz</th>
<th>Treatment Group</th>
<th>M</th>
<th>SD</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Individual</td>
<td>67.50</td>
<td>17.37</td>
<td>( t(16) = -5.07, p &lt; .05_a )</td>
<td></td>
</tr>
<tr>
<td>2 Individual Best</td>
<td>92.50</td>
<td>6.12</td>
<td>( t(16) = -1.02, p &lt; .05_b )</td>
<td></td>
</tr>
<tr>
<td>3 Cooperative</td>
<td>95.83</td>
<td>6.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Individual</td>
<td>96.25</td>
<td>1.02</td>
<td>( t(16) = 0.15, p = 0.88_a )</td>
<td></td>
</tr>
<tr>
<td>2 Individual Best</td>
<td>100.00</td>
<td>.00</td>
<td>( t(16) = 1.50, p = 0.15_b )</td>
<td></td>
</tr>
<tr>
<td>3 Cooperative</td>
<td>95.83</td>
<td>6.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Individual</td>
<td>82.24</td>
<td>16.77</td>
<td>( t(16) = -0.92, p = .37_a )</td>
<td></td>
</tr>
<tr>
<td>2 Individual Best</td>
<td>97.08</td>
<td>6.20</td>
<td>( t(16) = 2.06, p = .06_b )</td>
<td></td>
</tr>
<tr>
<td>3 Cooperative</td>
<td>89.17</td>
<td>10.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Individual</td>
<td>76.25</td>
<td>8.12</td>
<td>( t(16) = -1.92, p = .07_a )</td>
<td></td>
</tr>
<tr>
<td>2 Individual Best</td>
<td>91.67</td>
<td>8.62</td>
<td>( t(16) = 1.42, p = .18_b )</td>
<td></td>
</tr>
<tr>
<td>3 Cooperative</td>
<td>85.00</td>
<td>10.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Individual</td>
<td>76.25</td>
<td>13.89</td>
<td>( t(16) = -4.12, p &lt; .05_a )</td>
<td></td>
</tr>
<tr>
<td>2 Individual Best</td>
<td>94.17</td>
<td>8.75</td>
<td>( t(16) = -1.61, p = .13_b )</td>
<td></td>
</tr>
<tr>
<td>3 Cooperative</td>
<td>100.00</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Individual</td>
<td>50.31</td>
<td>11.35</td>
<td>( t(16) = 1.03, p = .32_a )</td>
<td></td>
</tr>
<tr>
<td>2 Individual Best</td>
<td>75.83</td>
<td>17.30</td>
<td>( t(16) = 3.42, p &lt; .05_b )</td>
<td></td>
</tr>
<tr>
<td>3 Cooperative</td>
<td>41.67</td>
<td>24.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Individual</td>
<td>56.25</td>
<td>8.77</td>
<td>( t(16) = -5.10, p &lt; .05_a )</td>
<td></td>
</tr>
<tr>
<td>2 Individual Best</td>
<td>73.33</td>
<td>8.17</td>
<td>( t(16) = -2.10, p = .05_b )</td>
<td></td>
</tr>
<tr>
<td>Quiz 11</td>
<td>1 Individual</td>
<td>66.35</td>
<td>10.23</td>
<td>$t(16) = -5.17, p &lt; .05_a$</td>
</tr>
<tr>
<td>Quiz 12</td>
<td>1 Individual</td>
<td>76.53</td>
<td>6.35</td>
<td>$t(16) = -3.34, p &lt; .05_a$</td>
</tr>
</tbody>
</table>

Note: $N = 66$; Quizzes 3, 4, and 5 were removed due to same testing methods and lack of variation; $t$-tests $a =$ comparison between individual and cooperative mean scores, $b =$ comparison between the best score of individual quiz scores and the best score of the cooperative quiz scores.

Statistical significance was set with an alpha level of .05 and was found in only a few of the quiz grade comparisons. Five of the nine quizzes had statistical significance between the scores. Quiz 1 had statistical significance between the overall quiz scores administered cooperatively and individually $t(16) = -5.07, p < .05$ and for the best scores of each group $t(16) = -1.02, p < .05$. Quiz 8 had statistical significance between the average of the overall quizzes scored individually and cooperatively $t(16)= -4.12, p < .05$. Quiz 10, 11, and 12 also had statistical significance ($t(16)= -5.10, p < .05$, $t(16)= -5.17, p < .05$, $t(16)= -3.34, p < .05$, respectively). Quiz 9 had statistical significance between the best scores of individual quiz scores and the best cooperative quiz scores ($t(16)= 3.42, p < .05$). In this case, however, the best individual score was higher than the cooperative quiz scores.

The cooperative quiz scores were higher than both the individual scores and the individual best scores on quizzes 1, 8, 10, 11, and 12. There was no statistical significance for best scores. Scores were higher for those weeks but not significantly so. The lack of significance could be due to ceiling effect which limited the potential effect for cooperative testing scores. Cooperative scores were higher than the individual scores, but not higher than the individual best scores for quizzes 6 and 7.

**Research question 2.** Does the use of cooperative testing increase student motivation toward science in a 9th grade biology course as measured by questionnaire responses?
Data taken from the pre-study, mid-study, and post-study student responses were collected to determine student motivation from the Glynn (2011) Science Motivation Questionnaire II (SMQ-II). Students were given five choices to answer questions on science motivation. Response selections were never (0), rarely (1), sometimes (2), often (3), and always (4). The overall possible score for motivation from the questionnaire was 0 – 100. Twenty-five questions were categorized by Glynn et al. (2011) into five subscales that support motivation for learning science. The mean item scores on the five subscales are broken down by related questions in Table 2. They are intrinsic motivation, career motivation, self-determination, self-efficacy, and grade motivation. Scores on each subscale could range from 0 to 4.

Table 2

Motivation Toward Science Questionnaire Responses

<table>
<thead>
<tr>
<th>Motivational category</th>
<th>Pre-study</th>
<th>Mid-study</th>
<th>Post-study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrinsic motivation</td>
<td>$M=2.72$</td>
<td>$2.61$</td>
<td>$2.90$</td>
</tr>
<tr>
<td></td>
<td>$SD=0.86$</td>
<td>$0.93$</td>
<td>$0.86$</td>
</tr>
<tr>
<td>Career motivation</td>
<td>$M=2.73$</td>
<td>$2.55$</td>
<td>$2.72$</td>
</tr>
<tr>
<td></td>
<td>$SD=1.03$</td>
<td>$1.05$</td>
<td>$0.89$</td>
</tr>
<tr>
<td>Self-determination</td>
<td>$M=2.71$</td>
<td>$2.57$</td>
<td>$2.70$</td>
</tr>
<tr>
<td></td>
<td>$SD=0.83$</td>
<td>$0.86$</td>
<td>$0.85$</td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>$M=3.11$</td>
<td>$3.09$</td>
<td>$3.25$</td>
</tr>
<tr>
<td></td>
<td>$SD=0.76$</td>
<td>$0.85$</td>
<td>$0.69$</td>
</tr>
<tr>
<td>Grade motivation</td>
<td>$M=3.33$</td>
<td>$3.35$</td>
<td>$3.40$</td>
</tr>
<tr>
<td></td>
<td>$SD=0.74$</td>
<td>$0.74$</td>
<td>$0.57$</td>
</tr>
</tbody>
</table>

Note: $N=66$
A one-way repeated measures ANOVA was conducted for each of the five subscales with the repeated factor being time (pre-study, mid-study, and post-study). The sphericity assumption of the ANOVA test was met with a non-significant Mauchly’s sphericity test. The results of the ANOVA for intrinsic motivation indicated a significant time effect, $F(2, 130) = 56.77, p < .05$. Follow up pairwise comparisons on the means in Table 2 using Bonferroni correction revealed that the results between the intrinsic motivation decreased on the mid-study from the pre-study, and was the highest on the post-study ($M = 2.90, SD = .86$).

A one-way repeated measures ANOVA was conducted for career motivation. The sphericity assumption of the ANOVA test was met with a non-significant Mauchly’s sphericity test. The results of the repeated measures ANOVA for career motivation indicated a significant time effect, $F(2, 130) = 24.11, p < .05$. Follow up pairwise comparisons on the mean scores using Bonferroni correction revealed that the time between the pre-study and mid-study results and between the mid-study and post-study were statistically significant ($p < .05$), however not statistically significant between the pre-study and the post-study for career motivation. Career motivation response mean decreased slightly in the mid-study questionnaire ($M = 2.54, SD = 1.04$). There was no statistically significant difference between the post-study and the pre-study.

A one-way repeated measures ANOVA was conducted for self-determination. The sphericity assumption of the ANOVA test was met with a non-significant Mauchly’s sphericity test for self-determination. The results of the repeated measures ANOVA for self-determination indicated a significant time effect, $F(2, 130) = 20.03, p < .05$. Follow up pairwise comparisons using Bonferroni correction revealed that the self-determination between the pre-study and mid-study results and between the mid-study and post-study were statistically significant ($p < .05$), however not statistically significant between the pre-study and the post-study for self-determination just as in the career motivation scale. Self-determination motivation response mean decreased slightly from pre-study ($M = 2.71, SD = .83$) to the mid-study ($M = 2.57, SD = .86$). The mean increased back to the pre-study level ($M = 2.71, SD = .83$) on the post-study assessment ($M = 2.70, SD = .85$).

The fourth category was for self-efficacy or how well the students believed in their ability to be successful in science. A one-way repeated measures ANOVA was conducted for self-efficacy. The sphericity assumption of the ANOVA test was met with a non-significant Mauchly’s sphericity test. The results of the repeated measures ANOVA for self-efficacy indicated a significant time effect, $F(2, 130) = 30.46, p < .05$. Follow up pairwise comparisons of the means from Table 4 using Bonferroni correction revealed that self-efficacy between the pre-study and mid-study results were not statistically significant; however, there were statistically significant ($p < .05$) differences between the pre-study and post-study and mid-study with post-study for self-efficacy. Post-study self-efficacy was higher than self-efficacy at pre or mid-study.
The fifth category was for grades as a motivating factor for science. A one-way repeated measures ANOVA was conducted for grades. The sphericity assumption of the ANOVA test was met with a significant Mauchly’s sphericity test, therefore Huynh-Feldt corrected significance level was used. The results of the repeated measures ANOVA for career motivation indicated no significant time effect, $F(1.42, 92.04) = 2.75, p = .09$. Follow up pairwise comparisons using Bonferroni correction revealed that grades as a motivational factor between the pre-study, mid-study, and post-study were not statistically significant. This scale ranked the highest of the five scales for motivation at pre-study ($M = 3.33, SD = .74$), mid-study ($M = 3.35, SD = .74$), and post-study ($M = 3.40, SD = .57$). Students scored high on this motivational scale throughout the study.

Overall, students in the research study were motivated toward science according to the questionnaire responses. Grades were the highest ranking subscale throughout the study. Self-efficacy and intrinsic motivation were the next highest ranking subscales with career motivation and self-determination being ranked least by the students. There was a slight decrease in the responses during the mid-study questionnaire.

**Research question 3.** Does the use of cooperative testing affect students’ study time in preparation for quizzes and final exam as measured by questionnaire responses?

The pre-study, mid-study, and post-study questionnaires included a question for students to answer about their anticipated or actual study time each week. Additionally, two questions were provided to students after the quiz each week asking about their study time amount. The percentages of the study time responses for students are based on 0-2 hours, 3-4 hours, or 4 or more hours.

The weekly questionnaires with open-ended responses for study time were “How many hours did you study for the quiz this week?” and “Do you think the amount of time you studied was sufficient? Why or why not?” The most common response (98%) was the length of the video for the week (i.e. a 7-minute video for the week, students responded they studied for “7 minutes”). Only one student responded on a weekly basis as studying for more than an hour each week.

**Qualitative analyses.** Qualitative data were collected in the pre-study, mid-study, and post-study questionnaires in order to provide depth to the students’ responses. The responses addressed such questions as their experiences with cooperative groups, advantages and disadvantages of testing in cooperative groups, group functioning, and group effectiveness. Additionally, responses were collected from 10 semi-structured, individual interviews and one focus group at the conclusion of the study.

Based on the responses, students had not worked previously in cooperative groups for testing, but had favorable attitudes for working in cooperative testing groups after their experience. The most common response was they could “learn from others” in cooperative testing groups and when they “could not remember some information, someone else in their group did.”
indicated positive reliance on group members and a healthy functioning of base groups. The most common drawback to working in the groups was they “sometimes disagreed.” Future benefits from the cooperative testing method were viewed as “better grades” and “team work.”

Several questions on the pre-study, mid-study, and post-study addressed student expectations, group functioning, and group effectiveness. The majority of the students felt positively about the group functioning well and about being able to work with those in their group. Overall, the students felt they could work well with others. The only notable difference was about being able to “give feedback to other group members about their understanding of their ideas.” The post-study revealed a decrease in the responses of the slightly agree to agree in the post-study questionnaire.

The responses indicated the majority would rather take the quiz in a group on the mid-study questionnaire and the post-study questionnaire. Before the study began, the responses were mixed between all of the responses. The pre-study, mid-study, and post-study questionnaire also included their experience in the past with cooperative testing. On the pre-study questionnaire, most all of the students indicated they had never had experience with cooperative testing, however responded “neutral” to this question because there was not an option to skip this question if they had not ever had experience with cooperative testing. The majority of the students felt “positive” to “strongly positive” for taking cooperative quizzes on the mid-study and post-study questionnaires. This was a major increase from the pre-study questionnaire responses indicating the same.

Additionally, student responses were collected for how much “fun” they thought the cooperative testing would be or was compared to individual testing. In the pre-study survey, 79% responded neutrally to the fun factor. By the post-study survey, 75% responded they thought taking quizzes in groups was “fun or a lot of fun.” Overall students had a positive experience and the main theme that emerged from their responses was “enjoyment.” Additionally, students noted their grades were better with the cooperative testing, which was supportive as a motivation toward science factor. They had fun working with their peers and felt the cooperative testing method would be applicable to other classes. Students felt the method would be applicable in certain classes or subjects.

4. DISCUSSION

Summary of research question 1. Academic achievement was examined through two sources. The first was the comparison of six cooperative group quiz scores versus the six quizzes taken by students individually. The second source for scores was the final semester exam questions to evaluate retention of the learned material. However, unbeknownst to the researcher until after the exam, the instructor for the course provided the students with a study
guide that had each of the exam questions verbatim. Therefore, the students’ scores on midterm were not valid for the research study.

The weekly quizzes consisted of 10 questions in various formats in increasing difficulty from multiple-choice questions to short paragraph response questions. The topics were based on the regularly outlined topics for the biology course adhering to curriculum standards such as characteristics of life, cell structures, DNA, Mendelian genetics, mitosis, and taxonomy. The first three weeks of the study and flipped lessons assigned consisted of review from biology lessons in prior science courses for students. Student scores were higher in all but two of the weeks of quizzes for the cooperative quiz method over the individual quiz scores.

Because the cooperative scores were higher than the individual scores in all but two of the quizzes, there was a need to investigate if this may be attributed to the highest scoring individual in the group instead of the cooperative effort between group members. The highest scores from the cooperative quizzes and the highest individual scores were compared for nine of the 12 weeks of testing through t-tests. Three of the weeks (3, 4, and 5) were the same format (individual or cooperative) and therefore not able to be compared. Results indicated that the higher scores were statistically significant for five of the quiz format comparisons.

In weeks 9 and 10, overall grades in both the cooperative and the individual quiz administrations were much lower than the other weeks of testing. This could have been attributed to the extracurricular activities in which all the students were involved. Many did not complete the homework assignment and stated they “forgot there was a quiz.” The material in week 9 was also material on taxonomy which is typically more difficult information for students to grasp and was new information for the students. Weeks 2 and 9 were the only two weeks in which the individual scores were higher than the cooperative scores. Week 2 had a difference of only 0.42 t-value between the individual average scores and the cooperative scores. This was a narrow difference and may be attributed to the information being review information with which the students were familiar.

Summary of research question 2. Motivation toward science for students was evaluated from student responses to the pre-study, mid-study, and post-study questionnaires including the Science Motivation Questionnaire II (SMQ II) (Glynn et al., 2011). It is probable that because the students in the study are in a college preparatory school, their motivation was generally high from the beginning of the study. Approximately 73% of the students responded they were “motivated to learn science” in the pre-study questionnaire. Even before they knew their group member assignments, 52% of the students responded they felt their “group would be motivated to learn science.” This indicated not only are they motivated to learn science, but also the majority of them felt their peers were as well.

The 25 statements of the SMQ II (Glynn et al., 2011) were separated into the five categories of motivating factors, as offered by Glynn et al. (2011). The categories were intrinsic motivation, self-determination, self-efficacy, career motivation, and grade motivation. According to the responses from the students on the questionnaire, grades were the highest
motivating factor followed by self-efficacy, career motivation, intrinsic motivation, and finally self-determination. Because grades were the highest factor in motivation for science and the quiz grades were higher on cooperative quizzes, it is understandable for the students to have favorable opinions on the cooperative testing method. Therefore, the academic achievement supported the motivating factor of grades. In four of the five categories (grades excluded), the mid-study response means decreased slightly and two remained lower for the post-study (career motivation and self-determination) while two increased above the pre-study responses for the post-study responses. This could be attributed to the fact that week 6 or the week before the mid-study questionnaire began the material that was newest for the students and this could have decreased their confidence in the lessons and their abilities. The motivating factor of grades, however, increased slightly over the pre-study, mid-study, and post-study questionnaires and remained the highest motivating factor of the five categories.

Interestingly, each question response in the category of self-efficacy increased from the pre-study questionnaire to the post-study questionnaire. This indicated student growth over the study period in their beliefs about their abilities. According to Alderman (2008), determining students’ self-efficacy is a way to not only better understand students with little motivation, but also student giftedness. Self-efficacy can motivate students for the tasks they will attempt and the amount of effort and persistence involved in the task.

**Summary of research question 3.** Study time data were collected from pre-study, mid-study, and post study questionnaires and from weekly questionnaires provided to students after taking their quizzes each week. The majority of students (98%) responded they only studied the length of the video for the lesson each week on the weekly questionnaires. Additionally, student responses to the pre-study questionnaire indicated 68% would study 0-2 hours each week for science and 28% would study 3-4 hours each week for science. Responses shifted to a much higher percentage of 0-2 hours study time for science each week on the mid-study questionnaire with 94% and 96% for 0-2 hours on the post-study questionnaire. This could be interpreted as the course was not as rigorous as the students thought it might be in the beginning of the school year.

A few students indicated they took 1-2 hours each week to study. Even though student notes on videos were not included as part of the study, notes were collected and checked for completion. Student scores on the quizzes corresponded with the amount of notes and quality of notes they took. It was apparent through the responses on the weekly questionnaires that if they did not take notes then they were also not pleased with their quiz performance on the “Do you think the amount of time you studied was sufficient?” Inevitably, if the students did not take notes, they indicated “0 minutes” on the amount of time taken to study for the quiz and also responded they did not do well on the quiz.

With the exception of week 9 when the group quiz score ($M = 42, SD = 24.83$) fell below the individual quiz score ($M = 50.31, SD = 11.35$), student responses were indicative they felt the amount of time studied was sufficient. Even though they only
studied the length of time of the video, they typically felt they were prepared sufficiently to do well on the quiz. Only utilizing the length of the video for study time indicated two things. The first was the students were not pausing the video long enough to take notes as directed in the initial instruction on flipped classroom lessons. Secondly, this could be an indicator that the lessons were too easy for the students that they were not using more time to prepare for the quizzes.

5. Qualitative analyses

A pre-study, mid-study, and post-study questionnaire was provided to students with questions pertaining to their experiences with cooperative testing, group functioning, their enjoyment related to the testing method, and group effectiveness. Semi-structured interviews with 10 students and one focus group of 10 students was also conducted to gather feedback.

Overall, students felt positively about cooperative testing and would like to continue the method for other courses. The most common response to the positive experiences with cooperative testing was the ability to have particular information from the video recalled by a group member during discussions for cooperative quizzes. The most common disadvantage to the cooperative testing method for quizzes according to responses was disagreement between group members when formulating answers for the quizzes.

The semi-structured interviews and the focus group also provided some insight to student perceptions about the cooperative quiz taking. There was discussion about the ability to use the method in other classes in the future and how this could be beneficial to them. One student remarked, “I think it depends on the subject.” Interestingly, a history teacher at the school was so intrigued by the cooperative quiz taking, he decided to try the method with his students. He reported positive results and feedback from his students. He also stated, “Students gained a greater amount of confidence from working with their peers and were more apt to contribute to discussions through the cooperative testing process.” At the conclusion of the study, the entire history department of the school was seeking professional development to assist in its efforts to use cooperative testing across a wider range of classes.

6. Conclusion of results

The statistical analysis of the results provided positive academic achievement results. Academic achievement on quizzes was much higher for cooperative quizzes than the individual quizzes for all but one week. The one week anomaly that had higher individual scores than cooperative scores was also a week of extracurricular activity and student responses to the weekly questionnaire indicated they had “not prepared sufficiently” for the quiz that week.
Motivation toward science in the research population of students was initially high and remained high for the duration of the study. In some areas of factors for motivation, such as self-efficacy, student responses increased collectively. Student enjoyment and group satisfaction rates were also high for the duration of the study for the cooperative testing setting. Positive support for the cooperative quiz taking was also indicated through the individual interviews as well as the focus group of 10 students. Positive responses for the need for cooperative work for their future success was a common thread in many responses.

In the pre-study questionnaire, more students responded they would study 3-4 hours (69%) than the student responses on the mid-study (5%) and post-study questionnaire (2%). These students dropped to the 0-2 hours of study time on the mid-study and post-study questionnaires. This indicated students studied less than originally thought. According to the weekly questionnaires which inquired about the length of study time for the students, 98% responded they studied only the length of the video for the week. Therefore, no significant difference in study amount was noted from the quiz taking methods. This may also be a result that the type of quiz taken each week was selected at random and not made known to them ahead of time.

7. Implications for instructional design

Instructional designers can also use the information for motivation toward science for high school students from this study. Motivated students engage in the material being learned and gain greater self confidence in their own abilities while learning more about the world around them (Glynn et al., 2011). Instructional designers may consider the positive benefits of students with cooperative testing and can find new ways of implementation. Cooperative learning and testing could be incorporated in a broader range of courses and subject areas. Additionally, cooperative testing could be implemented in younger age groups or grade levels. Increasing motivation toward science at even a younger age could help increase the number of students that would like to pursue science as a career choice.

This study also highlights the importance of empirical research for successful design of instruction for high school students. Being able to help students become more scientifically literate in today’s world is of utmost importance (Glynn et al., 2011). Data collected in this study showed increased academic achievement and motivation toward science for students. Therefore, instructional designers may consider these elements for continued research for cooperative testing in other settings for further empirical data collection and help increase our global standing for science education.

8. Practical significance

The practical significance of this study can be applied by instructors for students at any level. Students in this study saw value from cooperative testing and working with each other.
When students enjoy what they are doing or learning, they are more apt to spend more time on those tasks (Johnson et al., 2009). When motivation is increased, as it was in this study toward science, there is more time-on-task. Additionally, with more motivation comes an increase in confidence. Students even commented in their responses they were “more confident” when working in groups with the information learned.

Being able to help students relate to one another and be able to cooperate with one another is essential in our global world today. Classrooms have been using cooperative learning for many years, but have not taken the next obvious step to allow students opportunities to test in cooperative groups. Students view cooperative testing positively and see relevance for the future success in careers and the business world, as indicated from the responses on the questionnaires in this study. Instructional designers can create more job aids to assist any learning environment to incorporate these methods of cooperative testing and flipped classrooms.

Another practical significance of this study is that of cooperative testing being effective in large classes. Many of today’s classrooms have as many as 30 plus students in them. The workload of grading papers for teachers can be difficult with as many as 250 students per day. Therefore, cooperative testing can be a way to alleviate the time-consuming burden of test grading. This could allow teachers more time to help increase student learning.

With cooperative learning and testing, students have the ability to learn from each other. Sharing information learned with classmates can reinforce the material. Social interaction is a very important aspect of students’ lives and allows them the opportunity to learn and work together. This interaction can also help students function more successfully in social settings. Some students even commented they enjoyed working in groups that were selected for them because it forced them to work with students they did not know as well. The students in the groups began to rely on one another, which also increased their motivation toward science.

9. Strengths of the study

Having buy-in from all participants is extremely important in a study. The support of the course instructor was crucial. The initial work to set up the flipped classroom with the additional instruction time and the time to complete the weekly questionnaires was time taken from the instructor’s time for the class. Therefore, the study would not have been possible without the instructor’s support.

Classroom space can exhibit challenges in high school settings, however during this study, two adjoining rooms were available. When the students were in groups working cooperatively, they were able to be separated by the instructor and the researcher into two different, adjoining rooms. This enabled the students to be able to discuss the answers to the
tests during collaborative sessions without the risk of being overheard by other groups or disrupting other groups during the testing process.

Using a mixed methods research design with a treatment and control group and conducting a pre-study, mid-study and post-study questionnaire with focus groups was another strength for the study. There were a multitude of venues from which to collect the academic achievement data through the weekly test scores and final exam scores, and motivation and study time through the questionnaires. Being able to collect insight on the student perspectives was a key component in providing formative evaluation for adjusting the treatment and advancing the research.

Additionally, the questionnaire used for motivation data collection is the Science Motivation Questionnaire II (SMQ-II) created by Glynn, et al. (2011) provided valuable data. The indicators of motivation (intrinsic motivation, grade motivation, career motivation, self-determination, and self-efficacy) were developed empirically and helped to pinpoint student motivation toward science in this study.

10. Recommendations for future research

The possibilities for future research are vast for the topic areas in this study. First of all, this study could be replicated and have an extended study time length with a larger participant population. This study showed that most of these students at this age/grade level had opportunities to work in cooperative groups for labs and class activities, but few had ever taken cooperative quizzes. A year-long study would gather more data regarding how the students felt about testing in groups and how their academic achievement was impacted over a longer time period. The study could focus more on the interaction between the group members with a larger participant population.

Additionally, cooperative testing could also be evaluated in a different country to provide insight on how effective this technique is applied in different cultures. Or possibly, a study to research gender differences with academic achievement, motivation toward science, and study time for flipped classrooms with a cooperative testing design could be researched. Adult learners or higher ed settings could also be evaluated with cooperative testing and flipped learning.

In addition to a longer study period and increased number of participants, additional subject areas such as math, history, language or other elective classes could be included in a study. The flipped classroom technique with the cooperative testing could be used in other subjects or curriculum. This may provide additional data to compare with the results found in a science classroom.
11. CONCLUSION

This study helped to fill in gaps in research for the implementation of flipped learning and cooperative testing in secondary science classes. The findings showed academic achievement based on quiz scores was much higher on cooperative quizzes than on individual quizzes for all but one week of testing. However, statistical significance based on the $t$-test comparisons was only apparent in five of nine weeks of quizzes compared between highest scores on cooperative quizzes versus highest individual quiz scores. The $t$-tests did show the cooperative quiz scores were not statistically significant with the highest scoring individual quizzes. This supported the cooperative results between the students and the success not being indicative of the highest performing students in the groups.

Cooperative testing was shown to improve self-efficacy throughout the study on the questionnaire responses. According to Bandura (2001), efficacy is the center of self-regulation of motivation. Therefore, cooperative testing can improve student beliefs on what they can tackle and undertake. Motivation toward science for these students was high throughout the study based on pre-study, mid-study, and post-study questionnaires. There was only a slight increase in five of the factors of motivation as indicated by the Science Motivation Questionnaire – II (SMQ-II).

Study time was directly related to the length of the video. Through questionnaire responses, the majority of the students indicated they only spent as much time studying as the length of the video length per week. This indicated students were not utilizing full advantage of the flipped video design to be able to stop, pause, or rewind the video while taking notes. The effort in students’ notes was directly correlated with the outcome of their score on the quizzes whether it was cooperative or individual. Therefore, the more effort they put into their note-taking, the higher their quiz score.

Limitations were sample size, private school setting, and duration of the study. Strengths of the study included the instructor cooperation, the physical classroom setting, the motivation questionnaire, and mixed methods design for the study. Future research suggested conducting a longer study, utilizing a public school setting with a larger population size, conducting the study in different disciplines such as history or language classes, evaluating cultural differences for cooperative testing, and evaluating gender differences in cooperative testing methods. All of these could provide essential elements to the studies on cooperative testing in conjunction with flipped classroom lessons.

The practical significance of this study is that flipped lessons combined with cooperative testing can help increase academic achievement and motivation toward science. Not only can students benefit from these methods, teachers can reap the benefits as well from a reduction of papers to grade. There will be more time in class to focus on learning science, which can eventually lead to student proficiency in science and increase the U.S. student ranking in science education on the global scale.
12. REFERENCES


