

Chapter	Page
Statistical Analysis	25
4. RESULTS	29
Descriptive Statistics	29
Results of Research Question	43
5. CONCLUSIONS AND RECOMMENDATIONS	62
Overview of Findings	62
Limitations	63
Recommendations	64
REFERENCES	67
APPENDICES	76

LIST OF TABLES

Table	Page
1. Grading Rubric for Honors Physics Exams	19
2. Percent Composition of Honors Physics Exams	19
3. Honors Physics Standards	20
4. Example Grouping of Students for Homogeneous and Heterogeneous Grouping ..	21
5. Student Transition between Ability Groups	22
6. Ability Grouping Paired Sample Test between Pre-Force and Force Scores	47
7. Ability Grouping Paired Sample Test between Pre-Energy and Energy Scores	54

LIST OF FIGURES

Figure	Page
1. FCI Score Distribution	30
2. FCI Results Class Comparison	30
3. FCI Ability Results	31
4. EMCS Score Distribution	32
5. EMCS Class Results	33
6. Energy Concept Ability Results	34
7. Pre-Force Score Distribution Study Year	35
8. Pre-Force Score Distribution Comparison Year	35
9. Pre-Force Class Results	36
10. Pre-Force Ability Results	37
11. Force Group Productivity Survey Responses	38
12. Force Group Understanding Survey Responses	39
13. Force Group Repeat Survey Responses	40
14. Force Dyad Productivity Survey Responses	41
15. Force Dyad Understanding Survey Responses	42
16. Force Dyad Repeat Survey Responses	43
17. Force Summative Distribution Study Year	44
18. Force Summative Distribution Comparison Year	44
19. Force Class Results	45

Figure	Page
20. Force Ability Results	46
21. Pre-Energy Score Distribution Study Year	48
22. Pre-Energy Score Distribution Comparison Year	48
23. Pre-Energy Class Results	49
24. Pre-Energy Ability Results	50
25. Energy Summative Distribution Study Year	51
26. Energy Summative Distribution Comparison Year	51
27. Energy Class Results	52
28. Energy Ability Results	53
29. Energy Group Productivity Ability Survey Responses	56
30. Energy Group Understanding Ability Survey Responses	57
31. Energy Group Repeat Ability Survey Responses	58
32. Energy Dyad Productivity Ability Survey Responses	59
33. Energy Dyad Understanding Ability Survey Responses	60
34. Energy Dyad Repeat Ability Survey Responses	61

LIST OF APPENDICES

Appendix	Page
A. FORCE UNIT SUMMATIVE EXAM	76
B. ENERGY UNIT SUMMATIVE EXAM	80
C. CONTEXT-RICH PRACTICE PROBLEM EXAMPLE	84

CHAPTER 1

INTRODUCTION

The ability to work with others to complete a task is vital in almost every career, this statement is especially true for the scientific community. Large scale research projects require teams of scientists and engineers to collaborate together to shed light on the complex natural world. The Laser Interferometer Gravitational-Wave Observatory (LIGO) for example had a collaboration of more than 1000 scientists from 83 institutions worldwide working to measure gravitational waves (Chang, 2016). The Large Hadron Collider (LHC) at CERN is a collaboration that has created a team of some 10,000 physicists from around the world (Merali, 2010). Even an “isolated” scientist requires communication with the greater scientific community to test and validate their theories or experimental findings. Because of this increased pressure from the professional community, group work has become a high priority for education research.

Definition of Problem

Some in the education field have argued that all learning may derive exclusively from social interactions (e.g. Bandura, 1977, p. 192). Group work teaches interpersonal skills including: communication, conflict resolution, decision making, problem solving, and negotiation (Webb, Nemer, Chizhik & Sugrue, 1998, p. 611). These skills are critical to success in the professional physics community and should be a focus for physics classroom. Current

physics education research (PER) has encouraged a shift in pedagogy to a collaborative learning format. Cooperative learning occurs where instructors organize students into small groups which then work together towards learning the academic content (Slavin, 2011, p. 344). Some examples of such include Peer Instruction (Mazur, 1996), and Tutorials in Introductory Physics (McDermott & Schaffer, 2002). These methods also coincide with an increase in epistemology, also known as attitudes, beliefs, and views about knowledge and learning, regarding the nature of physics and learning physics in comparison to traditional instruction (Zhang, Ding, & Mazur, 2017). The effectiveness of collaborative learning is well documented but the organization of these groups is still being debated. There is no consensus on the most effective ability group composition, where evidence has been found in favor of both homogeneous and heterogeneous ability grouping methods (Jensen & Lawson, 2011).

Significance of the Study

This study will also investigate the arrangement of students by their academic achievement. Vygotsky argued that students have a potential development beyond their current ability level, known as a zone of proximal development and as summarized by Jensen and Lawson (2011) he “believed that children could perform above their current level of development when collaborating with others of higher ability (to a limited extent, of course)” (Vygotsky, 1978; Woolfolk, 2007). If students are helped by a fellow student with higher ability in physics they may gain the ability to use the physics concepts on their own (Ormrod, 2000; O'Donnell, Hmelo-Silver, & Erkens, 2006). To Vygotsky the teacher is a facilitator, while the students take the role of problem solvers. Thus the students become responsible for not only their own learning, but also that of the group (Bennett, 2015). Vygotsky's theory has been used to

justify heterogeneous ability grouping of students (e.g., Mugny & Doise, 1978; Tudge, n.d.) and is proposed to lead to a greater ability in scientific reasoning (Jensen & Lawson, 2011). The effectiveness of these ability based grouping methods has been a point of contention when concentrating on high ability students. Homogeneous dyads of high ability students are shown by some studies to outperform their heterogeneous dyads while working more collaboratively, engaging in more cognitive conflict and resolution, as well as producing better quality work (Hooper, Ward, Hannafin, & Clark, 1989; Fuchs, Fuchs, Hamlett, & Karns, 1998). However, another study found the opposite, this was due to the high ability students in the heterogeneous groups assumed the role of teacher and gave more explanations thus improving their understanding (Webb, 1980).

Working in groups and dyads can foster improved learning and comprehension (Beebe & Masterson, 2012, p. 13). Although many students report a feeling that they can accomplish the task individually better themselves than in a group, instructors typically find the group work helps students apply knowledge to a level not obtainable by an individual (Elgort, Smith, & Toland, 2008). Aside from teaching the value of teamwork, group work can help engage students in their studies and reinforce (not supplement direct instruction) content knowledge, assuming they stay on task (Bennett, 2015). Group work is shown to increase student reported satisfaction with the class. Simply adding diversity to the class helps student approval, in fact the more time students spend working in groups or dyads, the more favorable student attitudes on learning are reported (Springer, Stanne, & Donovan, 1999). This increased satisfaction encourages student learning and benefits larger classes especially in math and science (Lou, Abrami, Spence, Poulsen, Chambers, & D'apollonia, 1996, p. 451). Student self-esteem also is increased when group work is used for instruction, this is specifically favorable for female students who report

an increased feeling of competence (Springer et al., 1999). This more favorable attitude towards learning helps motivate students through STEM (Science, Engineering, Mathematics, and Technology) courses and programs at the university level to a greater degree than exclusive traditional methods of teaching (Springer et al., 1999). Finally the process of working in groups is highly valued by employers (Burke, 2011). Teamwork skills as well as social interactions are learned during group work, this may increase student exposure to various backgrounds, cultures, beliefs, and attitudes (Burke, 2011; Lou et al., 1996, p. 425).

Research Questions

To develop an understanding of the composition of collaborative learning groups and dyads in the high school physics classroom, the data analysis and summary of results were based on the following research question:

Does heterogeneous or homogeneous academic ability groups and dyads have an effect on the academic achievement of students in the high school physics classroom?

Limitations of the Study

This study will investigate the use of group and dyad work for a high school junior level honors physics course of 76 students. This sample is small and results will not be generalized for a larger population. Students will participate via completion of summative exams and reflective survey questions. The exams will be compared to previous year's student scores to help identify any significance; though these exams are not identical, their differences will be discussed. The survey questions will remain constant throughout the study however they will initially be taken anonymously.

CHAPTER 2

LITERATURE REVIEW

Definition of Groups and Dyads

The general definition of group work refers to the physical placement of students into groups for the purpose of learning (Lou et al., 1996, p. 423). Firstly we must discuss the number of members. According to Beebe and Masterson, a dyad is a partnership of two members while a group of at least three adds a complexity in communication that can pit two members against the third (Beebe & Masterson, 2012, p. 5). The larger the number of group members inversely relates to the potential influence of each group member while simultaneously creating more individual member interactions, which may cause subgroups to form (Beebe & Masterson, 2012, p. 6). This implies that there is an upper limit on student group size, however there is the limitations of the total class size and materials in a given course to take into consideration. Heller, Keith, and Anderson (1992) noted that the optimal group size in physics classes was three members, at this point the group was large enough to create multiple ideas regarding the physics concept and approaches to solving physics problems with each member having a voice, yet small enough to be manageable (Heller et al., 1992). Others have suggested groups of four to attempt to avoid having one student treated as an outcast (Csernica, Hanyka, Hyde, Shooter, Toole, & Vigeant, 2002). It has also been noted that the shorter the amount of time the groups have to work together the smaller the number of member should be to allow for an appropriate exchange of ideas between members (Cooper, 1990; Johnson, Johnson, & Smith, 1991). Due to limitations in materials and class time as well as total number of students our study will concentrate on groups of four for both practice problems as well as laboratory work in the subject of physics.

Dyads are not recommended to engage in activities that are designed for group work, because there are an insufficient number of students to generate creativity and a diversity of ideas (Csernica et al., 2002). This may result in members of dyads functioning as a learner and a teacher, which essentially creates a learning group of one (Lou et al., 1996, p. 430). However, prior studies have shown that dyad training is as effective as individual training (Day, Arthur, & Shebilske, 1997; Shebilske, Regian, Arthur, & Jordan, 1992), yet these studies did not report the quality of the interaction between the students (Crook & Beier, 2010, p. 337). The ideal result of dyad training would be the two trainees working together to learn a new task, knowing that the task will be assessed on an individual basis (Crook & Beier, 2010, p. 335). This type of learning differs from group work, in that the members should be able to complete the entirety of the task, as opposed to a group where tasks may be divided up and completion of the task requires all members to work together (see Kirschner, Paas, & Kirschner, 2009). To account for these limitations in dyad learning our study concentrates on dyad performance after the physics concept has been properly modeled by the instructor via the gradual release method (e.g. Coyle, Newman, & Connor, n.d.). More specific for the physics classroom this form of instruction is called the modeling method for physics (Wells, Hestenes, & Swackhamer, 1995).

Instructors have a variety of options in selecting members of groups, the descriptions below are an adaptation and combination from the work of both Linda Shalaway (2005) and Michael F. Opitz (1998). It should also be noted that there are many factors other than academic concerns discussed below when constructing groups, including gender, interest, ethnicity, and so on. This study will investigate the academic achievements of the students as it pertains to group and dyad arrangements in a physics classroom. The types of groupings for the four member groups and dyads are described in the following paragraphs.

Random grouping occurs when the instructor assigns members to groups unsystematically. This technique is often used for classroom management purposes mainly to control group size, which is especially useful for large classes (Davis, 2007, p. 151). This method requires the smallest amount of time and effort by the instructor while still allowing grouping to be controlled by the instructor.

Homogeneous grouping occurs when the instructor places students with similar achievement levels in the same group. These groups can be used to target specific skills that are needed for academic success, and allowing the instructor more time with those that are struggling with the class material. Homogeneous grouping can allow for high achieving students to challenge each other and hopefully achieve even greater success, while simultaneously allowing those that may be struggling to commiserate and work towards academic success. Homogeneous groups requires the most time and preparation by the instructor when compared to the other methods of grouping students as groups are arranged by ability level. Preparation is required for creating the task as Kulik and Kulik (1991) assert that ability grouping is pointless without instructor adaptation of the course material to the level of the students in each group.

Heterogeneous grouping occurs when the instructor intentionally separates achievement levels to allow for each group to have a mix of high and low achievers. This method can be used to have students help teach each other which theoretically benefits both the high and low achieving students (Webb, 1980). Heterogeneous grouping requires more time to create groups than the random method, however the course material does not need to be adjusted for each group.

The final grouping strategy to be tested in this experiment is student choice grouping, which allows the students to assemble the groups themselves. This method involves the least

amount of instructor preparation, however typically results in friends forming groups which results in more time off task (Cooper, 1990). Not surprisingly research shows student choice groups are lower performing in comparison to groups assigned by the instructor (Felder & Brent, 2001). Interestingly, research also shows that students reported their worst group work experience in student choice grouping, while reporting their best experiences occur in groups assigned by the instructor (Beebe & Masterson, 2012, p. 60). Instructor constructed groups help avoid outcasts, while teaching tolerance and strategies for working in groups (Lou et al., 1996, p. 427).

A term often used in education when referring to small groups is cooperative grouping, this type of group occurs when the instructor assigns each group member a different role. We have not classified this as a grouping method as it does not account for how the students are selected to join a specific group, but rather deals with the basic mechanics of how a group functions. Cooperative grouping does afford each student the chance at holding different positions in the group such as the leader or presenter or other specific roles. However, each of the previous methods for constructing a group (Random, Homogeneous, Heterogeneous, or Student Choice) can be made into a cooperative group by the assigning of roles by the instructor.

How to use Groups in Class

Barbara Gross Davis (2007 p. 147) adapted research to name three general types of group work as: informal learning groups, formal learning groups, and study teams. Informal learning groups are temporary groups usually existing for only a single class period to complete a learning task. Formal learning groups are teams built to complete a specific task such as labs or projects. Study teams are supportive groups that work together for an entire semester to complete

course assignments. (Davis, 2007, p. 148 adapted from Johnson et al., 1991). The first part of this study will investigate informal learning groups as the instructor will form student groups and dyads in multiple different arrangements to test the effect on student learning outcomes during a unit on force. The second part of this study will investigate formal learning groups as students will be paired with the same group for the entirety of an academic unit on energy.

There are four stages the instructor must complete to successfully integrate group work into the classroom as stated by Alison Burke (Burke, 2011). The first stage is for the instructor to decide how they wish to use group work and for what length of time, be it a portion of the class period, a portion of the semester, or for select activities. In planning group work the instructor must make sure that it is not seen as busywork and is directly tied to subject learning objectives. These tasks should also require interdependence so that the students know each individual is responsible for the success of the entire group (Davis, 2007, p. 149). To give purpose to each student's work, formative assignments will be spot checked at either the end of class or the beginning of the next class meeting.

The second stage to creating successful groups is to train the students how to communicate within a group, "the instructor cannot assume that students know how to work together, structure time, or delegate tasks" (Burke, 2011). The skills that should be discussed and modeled for the benefit of the small groups include giving and receiving constructive criticism, mastering content knowledge as a group, and managing disagreements in an appropriate manner (Davis, 2007, p. 148). A successful group should function as a team, one with clearly defined goals, members that understand and execute on their well-defined roles, and a clearly established procedure for communication (Beebe & Masterson, 2012, p. 6). This will be addressed for the

research study via instructor lead discussion at the beginning of the semester before the data analysis commences.

The third stage is to monitor the group performance. The instructor should help each group plan a procedure that includes who will be doing what and when. They then need to regularly check with each group to ensure they are moving towards their goal, as well as to aide in uncooperative group members (Davis, 2007, p. 152). It is also suggested that instructors may also want to assign roles in the group to decrease the number of problems (Heller & Hollabaugh, 1992). For this study the instructor will monitor group behavior during class and tend to difficulties during class time.

Finally, it is suggested that the instructor must decide and provide how group members will be assessed. Some instructors assess students individually, combining test scores as well as peer evaluations within the group. Other instructors may assign the same grade for each member, fearing competition within the group would be detrimental to group performance. Finally, another recommendation would be not to count the group work for too much, or any, of the overall student grade (Davis, 2007, p. 152). It is not how smart individual team members are, but how well they communicate that improves teamwork (Beebe & Masterson, 2012, p. 7). For the purposes of this study the instructor will not assess a summative grade, one that impacts the student's overall course grade, on group work assignments.

In some cases groups will not work well together, either due to a lack of motivation, poor leadership, or personality conflicts (Burke, 2011). Some students will fear group work, known as "grouphate". This feeling can be reduced when the group members receive proper training on how to communicate (Beebe & Masterson, 2012, p. 2). It also should be noted that the physical arrangement of students does affect the performance of the group, when students face one

another they tend to work better together, as opposed to being side by side (Heller & Hollabaugh, 1992). Regardless of group dynamics, the groups should be kept together until the completion of all work. Not only would problematic groups not learn how to cope with unproductive interactions, but the dispersion of this group into others may disturb their new groups (Davis, 2007, p. 151). For this study groups remained together for the predetermined time.

Group Work vs. Other Forms of Instruction

According to Anderson (1982) there are three phases of learning a skill. The first phase requires the learning to gain the basic knowledge about the skill, called the declarative stage. This phase results in a step-by-step process. The second phase is the knowledge compilation phase, where the learner assembles the steps of the process together into larger units. After sufficient practice the learner will enter the procedural knowledge phase, where performance becomes routine (Crook & Beier, 2010, p. 336). This study will primarily investigate the later stages of Anderson's theory where students will begin their group or dyad work at the beginning of the knowledge compilation stage.

Students will initially enter the declarative stage with whole-class instruction, or when all students are taught in a large group. Whole-class instruction allows the instructor to place an emphasis on uniformity of instruction, explanations, and encouragement (Lou et al., 1996, p. 424). Reasons for using whole-class instruction including uniformity of instruction where the instructor can concentrate on appropriate pedagogical measures for the introduction to a new skill, followed by whole-class practice to improve skills. Whole-class instruction also lends itself to instructors being able to emphasize a specific objective or skill for all students. This method

also provides a chance for competition among students which is more in line with the survival of the fittest system that dominates the professional atmosphere (Lou et al., 1996, p. 425).

After the students have shown comfort with the new material and begin to enter the knowledge compilation phase there are many advantages to small group work in the classroom. There is evidence that students in collaborative groups that provide explanations to teammates and discuss decision making steps lead to better learning outcomes (Dillenbourg, Baker, Blaye, & O'Malley, 1996). These groups have more information than any single member and can approach problems from a position with more potential creative solutions (Beebe & Masterson, p. 13). It is found that better physics problem solutions are formed via group work rather than individual work which has led to higher ability among students (Heller et al., 1992). If the problem is of a higher order, one in which no one student can answer individually, the chances of success for the group increased in comparison to individual work, in fact group work may have helped equalize resources among students (Webb et al., 1998). In addition, the instructor has greater flexibility in adjusting the objectives for each group depending on their learning needs, which can increase pace and rigger for the advanced students and allow the instructor to have more time with those struggling with the material (Lou et al., 1996, p. 425).

Group and Dyad Method Benefits and Shortcomings

Group and dyad work is ideal to add diversity to instruction, increase student's teamwork skills, and improve student attitude of the course. Students should be trained in working in groups or dyads, and should only work in groups or dyads when the material lends itself (Burke, 2011). Group instruction also shows greater return on learning outcomes when compared to whole-class instruction for K-12 education (Lou et al., 1996, p. 424, Jensen & Lawson, 2011).

Yet, these studies effect sizes only ranged from +0.17 to +0.32, which is in the small range for effect size. These studies were also limited by the size of the sample with only 15 studies for the meta-analysis (Kulik & Kulik, 1987, 1991; Slavin, 1987). In addition these studies only investigated group versus no group, and did not investigate the effects of homogeneous or heterogeneous student ability grouping (Lou et al., 1996, p. 424). Collaborative learning is an effective method to improve student learning, as well as retain STEM (science, technology, engineering, and mathematics) students (Rutherford & Ahlgren, 1990; Forman, 1989; Drew, 1996; Slavin, 2011; Lord, 1997; O'Donnell & King, 2009). In a calculus based college physics course for non-physics majors in China, Peer Instruction is shown to improve students' attitude and beliefs regarding the nature of physics as well as learning physics when compared to traditional college instruction (Zhang et al., 2017).

Multiple studies report that high achieving students in homogeneous grouping outperform their high achieving counterparts in heterogeneous groups (Webb et al., 1998; Hooper & Hannafin, 1988; Fuchs et al., 1998). In another study the high achieving students perform well in any group arrangement, yet the average achieving students benefit from homogeneous grouping (Lou et al., 1996, p. 449). Among elementary students homogeneous group members assisted each other more and achieved a higher achievement than heterogeneous group members where lower ability students would often allow the higher ability student to do all the work (Hooper, 1992; Fuchs et al., 1998). Webb (1991) found medium ability high school students to achieve more in homogenous groups. In science, Lawrenz and Munch (1985) found homogeneous groups to have larger gains in reasoning and better physical science achievements, yet a better predictor was still initial ability. For algebra homogenous groups encouraged teamwork, while heterogeneous groups lead to a follow the leader result where most students just copied (Weld,

1999). According to Lawson (1992) the ability of the partner plays an important role in motivation and attitude, noting medium ability students preferred working with similar or higher ability students. All students, especially low ability students, in an active inquiry-based college biology class were measured to performed better when in homogeneous groups (Jensen & Lawson, 2011). Harlow, Harrison, and Meyertholen (2016) measured a small difference for the freshman college physics classroom that there is an increase in student satisfaction with the lab portion of the course when sorted into homogeneous groups compared to heterogeneous groups. They noted that nine students in heterogeneous groups complained about group dynamics in contrast to no students complaining for homogeneous arrangements (Harlow et al., 2016).

Other researchers have measured a benefit to heterogeneous groups, resulting in more creative solutions to problems and increase in discussion as well as greater group satisfaction when compared to homogeneous groups (Amaria, Biran, & Leith, 1969; Bracey, 1994; Simsek, 1992; Webb, Nemer, & Zuniga, 2002). Lower ability students were more task oriented and even experienced a greater achievement in heterogeneous grouping (Bracey, 1994; Carter & Jones, 1994). For low level questions on Bloom's Taxonomy, gains were seen for heterogeneous groups in the traditional college biology classroom, which is consistent with other research (Jensen & Lawson, 2011; Webb et al., 2002). Thus, if the assessment concentrates on memorization rather than higher order questions regarding conceptual understanding heterogeneous groups show gains (Jensen & Lawson, 2011). The effectiveness of heterogeneous grouping in a mathematics course shows an improvement for the lower and average achieving students with no effect on the higher achieving students in comparison to homogeneous grouping (Linchevski & Kutscher, 1998). Lower achieving students benefit from heterogeneous groups due to access to high level solutions and quality explanations (Webb et al., 1998). Noddings (1989) suggests that in a

heterogeneous group the high achieving student may become a crutch for the other group members, thus diminishing communication within the group which may affect the attitude of its members.

No difference between heterogeneous and homogenous groups was measured for college students by Watson and Marshall (1995a, 1995b), though they predicted that heterogeneous groupings would result in a larger learning achievement due to greater student interaction as theorized by Vygotsky. Jensen and Lawson (2011) found no difference in achievement for medium and high ability college biology students regardless of the ability grouping implored. Harlow et al. (2016) found that there was no measurable effect a freshmen college physics student learning regardless of the ability makeup of the students in the groups.

CHAPTER 3

METHODOLOGY

This study's purpose is to determine effective group and dyad arrangements in the high school physics classroom in terms of academic ability. This chapter discusses the design and methodology of the study including the data collection and statistical analysis methods. The study was two part, the first part consisted of an overview of possible group arrangements on a short time scale. This was done with both dyads and groups of four. The purpose of the first part was to compare the group organization options and to allow for normalization of the data so comparisons can be made both within the first part and the second part. The second part of the study consisted of a classes having specific group organization for the entire unit. The purpose of the second part of the study was to collect a more complete set of data regarding subgroups as well as to provide a more complete comparison between organizational methods. Before data collection had commenced a class discussion was held regarding the expectations of group performance, this consisted of both instructor expectations and a student led discussion on expectations for peers.

Description of Sample and Population

The study investigates student performance in a junior level honors physics course at a high school in a large suburban school district during the fall semester of the 2015-16 school year. This sample demographics will be discussed in comparison to the school as a whole and compared to the state.

This study investigated a total of 76 students across three honors physics classes taught by the same instructor, to eliminate any instructor effect. In this case instructor effect refers to the natural differences, personality, experience, etc., between different instructors. The average class size was 25.3 students, which is large when compared to the school and state averages which were only 21 student per class. There were 41 female and 35 male students for this study. The previous year's class of 46 student across two honors physics classes with an average of 23 students in each will be used for comparison to notice any significance. There were 18 female and 28 male students for the previous year.

According to the school report card the High School's population at the time of this study is predominantly White followed by Hispanic. The school also has 36.3% low income student population (Illinois Report Card, n.d.). This study will not concentrate on ethnicity or income level, this is merely presented as a representation of the population.

The high school has an average ACT score of 20.7 with 50.6% of the students ready for college coursework. The ACT score is comparable to the state average, only 0.2% better, however the percentage of students that are ready for college coursework is 5% greater for the high school when compared to the state (Illinois Report Card, n.d.).

The specific academic performance of the students in the study will not be discussed in this paper due to limitations in collection of data as well as the privacy of students. However, it should be noted that the students of this study are all enrolled in honors physics, which is typically taken as a college preparation course typically taken during the junior year of high school and thus has a higher enrollment of academically high achieving students. The sample of physics students is not reflective of all students enrolled in high school physics, as they are in an

honors level course. However, these students are of the same age bracket which implies that the findings of this study may provide insight to the greater high school physics student population.

Research Question

Does heterogeneous or homogeneous academic ability groups and dyads have an effect on the academic achievement of students in the high school physics classroom?

Analysis of this question is based on both academic performance on concept tests and physics course unit exams, as well as student survey responses. These measures were then analyzed based on student academic ability when tiered into three ability levels.

Instrumentation

The first part of the study investigated student performance for the force unit of the honors physics classroom with diverse grouping options. The study stretched from the seventh week through the thirteenth week of the fall semester. Student ability was initially assessed in two ways and combined to get an overall ability ranking. Students were given the force concept inventory (FCI) as a pre-test at the beginning of the seventh week of the school year before they had any instruction on mechanics (Hestenes, Wells, & Swackhamer, 1992). The FCI was only used as a means of assessing student comfort and ability with physics for grouping purposes, it was not used as a post-test and did not affect student grades. This measured their prior understanding with both motion and Newton's Laws. The second quantity measured towards the student ability was their current grade. The current grade consisted of two examinations, covering one dimensional kinematics and two dimensional kinematics. These summative exams were taken at the end of week four and six respectively. The exams were scored with a rubric for

each question, the scores are described in general terms in Table 1. The current grade was then calculated as a weighted average between the questions.

Table 1
Grading Rubric for Honors Physics Exams

Score	Description
4	Mastery
3	Proficient
2	Basic Understanding
1	Below Basic Understanding
0	No Evidence

The exams were comprised of multiple types of questions including conceptual, numerical, and graphical based questions. The exams between years were different, yet highlighted the same material. The percent of each type of question is shown in Table 2 for all exams during this study and from the previous year. The force and energy unit exams for the year of study is located in Appendix A and Appendix B respectively for reference. The previous year's students' scores will be used for comparison.

Table 2
Percent Composition of Honors Physics Exams

Year	Question Type	Exam Name			
		1D Kinematics	2D Kinematics	Force	Energy
2015-16	Numeric	50	80	58	75
	Conceptual	20	20	42	25
	Graphical	30	0	0	0
2014-15	Numeric	54	79	58	73
	Conceptual	15	11	42	27
	Graphical	31	0	0	0

The standards for the study are given in Table 3 below, they are identical for both the year of this study and the previous year.

Table 3

Honors Physics Standards

Unit	Standard	Description
Kinematics	A1a	Kinematics Graphs
	A1b	Constant Velocity
	A1c	Constant Acceleration
	A2a	Vectors
	A2b	2D Projectile Motion
Force	B1	Newton's First Law
	B2	Newton's Second Law
	B3	Applications of Newton's Laws
Energy	C1	Work
	C2	Kinetic Energy
	C3	Potential Energy
	C4a	Conservative Forces
	C4b	Non-Conservative Forces
	C5	Power

Students were then ranked with the top score being assigned a value of one. These ranks were assigned for both the FCI and the current course grade. The overall ability of each student per class was then calculated by taking the average of the student's rank on the FCI and course work. The top third of this ranking is given the moniker of “high ability”, where the middle third is “medium ability”, and finally the bottom third is given the name “low ability.”

Student groups were then formed based off of their overall ability as discussed above. For dyads, homogeneous dyads simply paired off in order of ranking. While heterogeneous dyads took the top student and paired them with the lowest ranked, similar to a sports bracket. Identical methods were used for groups of four. An example of the method used to construct groups is given in Table 4, where students are first ranked then divided by the method described above.

Table 4

Example Grouping of Students for Homogeneous and Heterogeneous Grouping

Student	Rank	Homogeneous		Heterogeneous	
		Dyad	Group	Dyad	Group
A	1				
B	2	A,B	A,B,C,D	A,H	A,H,B,G
C	3	C,D	E,F,G,H	B,G	C,F,D,E
D	4	E,F		C,F	
E	5	G,H		D,E	
F	6				
G	7				
H	8				

For the second portion of the study, student performance was measured for the energy unit with recalculated groups and dyads using identical criteria after the first part of the study was completed. The energy unit ran from the tenth week to the twelfth week of the fall semester. The overall grade of the students was calculated by taking a weighted average of their scores on the kinematics and Newton's law standards. The kinematics standard carried a weight of 22% of their overall grade while Newton's laws carried a weight of 20% for their overall grade, the remaining percent is for other standards including energy, momentum, rotation, and gravitation. The weights of these standards were agreed upon by the physics curriculum team at the High School. The concept test used for student ranking for the second part was the energy and momentum conceptual survey (EMCS), which is a research-based test to assess student understanding of concepts in both energy and momentum units (Singh & Rosengrant, 2003). Similar to the FCI, this was used only as an initial indicator of student background understanding, so no post-test was given.

The overall student grade as well as their scores on the energy and momentum conceptual survey once again were ranked individually by class. These scores were averaged to get the

students' overall ability ranking. Resulting in new ability groups, though most students remained in their original ability group, there was some movement between abilities, see Table 5. Most notably are two students that moved from low to high ability and the one student that did the reverse as a result of the force exam.

Table 5
Student Transition between Ability Groups

Force Unit	Energy Unit		
	High	Medium	Low
High	19	6	1
Medium	4	12	9
Low	2	7	16

The ability ranking was used again by the instructor to create new groups in a similar method detailed above. However for this portion of the study each class was grouped in a separate configuration, the first period class was assigned partners and groups randomly, while the second period class was assigned with homogenous groups, and finally the last class period was assigned with heterogeneous groups. Comparison of each class are given in the results section of this paper.

Group and dyad work sessions alternate daily throughout the weeks of the study, after the sessions students were given a short survey requesting information based on how they felt the group had performed. Due to the pace of the high school setting the survey consisted only of three questions with a section for comments. The first two questions were on a Likert scale scored between one and five with five being favorable. The final question was a yes/no question. The questions were:

1. Was your group productive?
2. Do you feel you have a better understanding of the material?
3. Would you like to work in this group again?

For the first part of the study surveys were given anonymously. This anonymity in the surveys does present a problem of report subgroups responses. This was done to note student interest in grouping configurations and to set a baseline for comparison during the energy survey. The survey for the second part of the study was identical, yet students signed their surveys so subgroups could be discussed.

Data Collection

Data was collected during the fall semester of the 2015-16 school year for three honors physics classes. The honors physics class is a 50 minute daily class Monday through Friday. The data collection started in week 7 and ended in week 13, covering the duration of the Newton's Laws (Force) unit and Energy unit in the honors physics classes. A total of 76 students participated in the study with one common instructor.

Concept tests were taken at the beginning of each unit, week 7 for the force unit and week 11 for the energy unit. The FCI was utilized for the force unit and the EMCS was utilized for the energy unit. Unit exams were taken at the end of each unit, week 10 for the force and week 13 for the energy unit.

Group and dyad work was each given about twice a week during the study year, these activities were typically given after formal instruction by the instructor. For the previous year, the one used for comparison, group and dyad work was given less consistently. Dyad work occurred about twice a week in a less structured format by comparison to the year studied. Group

work given at most once a week, typically for labs. For the force unit survey responses were taken at the end of each class that involved dyad or group work. This occurred frequently, with the use of the modeling method of physics the instructor taught a new skill at the beginning of the class, then went over the material as a class, finally resulting in group or dyad work time. For the energy unit, surveys were given on the last day of the dyad or group meeting before the energy unit exam.

For dyad work, students were typically asked to work through physics practice problems ranging from conceptual questions to numerical application questions. On occasion students were asked to work on context-rich physics questions, which do not have an easy to obtain solution, and are intended to require teamwork to complete due to their difficulty (Heller et al., 1992; Heller & Hollabaugh, 1992). An example context-rich physics question that was used during the study is given in Appendix C and is adapted from the University of Minnesota (Physics Education Research and Development Group, n.d.). These assignments were not formally assessed by the instructor as they were meant as practice with the material obtained during direct instruction and were only checked by instructor observation.

For group work, students were asked to perform lab work or work on context-rich physics questions. During lab situations students were given the task to organize a structure for collecting data, then complete proper analysis of the data to obtain a reliable result. These labs were then turned in to the instructor for formative grading. Formative grading refers to the practice of not counting the assignment towards the overall grade of the student, but rather for the instructor to concentrate on comments to improve students' performance and understanding of the material. The context-rich physics questions were also formative and given to evaluate

student understanding of the mathematical nature of physics and apply the concepts to real-world situations.

Statistical Analysis

Some analysis and all graphical representations of both the exam and survey data will be done with Microsoft excel, while hypothesis testing will be done with IBM SPSS.

Student exam scores and current grades will be summarized using histograms and box plots. Histograms will count the number of students to achieve a score within a range (bin), they will give the reader an opportunity to see the distribution of scores. For box plots the median, the value lying at the midpoint of the data, will be plotted as a line of a box plot, with the 25th and 75th percentiles as the lower and upper endpoints of the box (Devore, 2004, p. 40). Thus the box will represent the middle half of the data. The “whiskers” will be lines that extend to the smallest or largest observation, but no further than 1.5 interquartile ranges from the 25th or 75th percentiles. The interquartile range is the difference between 25th and 75th percentile values. If any values exceed 1.5 interquartile ranges either above or below, they are considered to be outliers and will be represented by an empty circle point. In addition the box plot will also contain the mean of the data, represented by a solid point with label. Box plots will be used to compare different ability subgroups to each other and the total sample, using the raw exam and concept test data.

Diverging stacked bar charts will be used to represent simple number and percent comparisons for survey question responses. These bar charts will center the Likert ranking of three at the center and add the other rankings on either side to allow for easy comparison between groups.

Average normalized gain, $\langle g \rangle$, was introduced by Hake (1998) and is a rough measure of student learning. For the purposes of this study the average normalized gain will be altered as it will instead look at the average gain between unit exams, the calculation comparing pre-force and force average exam grades would be:

$$\langle g \rangle = \frac{\langle force \rangle - \langle preforce \rangle}{4 - \langle preforce \rangle}$$

Where the brackets represent the average of the enclosed exam or grade, and the denominator has a 4 for the maximum that is attainable on the exam. The sign of the average gain shows the direction of change.

For hypothesis testing both paired and unpaired t -tests will be used, which assumes that our sample is of a Gaussian distribution about the mean. According to Devore (2004, p. 300) a paired t -test will be used to compare the mean to two matched groups, whereas an unpaired t -test will be used to compare two unmatched groups. The t -test will be used to calculate the probability that two sample sets of data are statistically equivalent. The value of a t -test will be reported with the degrees of freedom in parenthesis next to the variable t .

When multiple groups need to be compared for statistical analysis, analysis of variance (ANOVA) will be used (Devore, 2004, p. 441). This method of statistical analysis can be implemented when groups have a similar variance. The result of the ANOVA analysis is referred to as F statistic, which is a measure of the variability in the scores between the conditions compared. The F statistic will be reported with the degrees of freedom for the variable and error in parenthesis.

The null hypothesis for these comparisons will typically be that the two samples have the same mean and any observed discrepancy between the sample means is due to random variation.

A two-tailed significance (p) value will be calculated as we do not know the direction of the change in mean and this method is more conservative than the one-tailed p value. If the p value is less than 0.05 then the two sets of data are considered different, thus rejecting the null hypothesis (Devore, 2004, p. 349). The standard 95% confidence interval (CI) will be used for reporting testing as well, it shows the range where there is a 95% chance that the 95% CI contains the true population mean (Devore, 2004, p. 281).

The effect size is a quantitative measure of the strength of a phenomenon, or a measure of how much the dependent variable is effected by the independent variable. When paired with the p value, more confidence of a correlation is presented. In this study we will use two measures of effect size. For a t -test we will be looking at the effect sizes based on differences between means, specifically calculating Cohen's d as introduced by Cohen (2009):

$$d = \frac{\langle x_1 \rangle - \langle x_2 \rangle}{s}$$

Where the difference between the averages is in the numerator, and the pooled standard deviation (s) is represented by:

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 + 2}}$$

Where s_1 and s_2 are the variance for the groups, and n_1 and n_2 are the number of samples per group. The results of this will be a simple numeric value, a value greater than 0.8 is considered large, a value between 0.2 and 0.8 is medium, and any value less than 0.2 is considered small (Cohen, 2009). These value will be used to compare similar tests throughout this paper.

For ANOVA testing we will use Eta squared (η^2) as our measure of the effect size, and is shown below:

$$\eta^2 = \frac{\text{sum of squares for the effect}}{\text{total sum of squares}}$$

This can be interpreted in a similar fashion to Cohen's d , however the bounds for a large effect would be greater than 0.14, a medium effect is 0.06, and a small effect would be less than 0.01 (Cohen, 2009).

In addition, the lower bound estimate of the reliability, or internal consistency, of the exams and survey will be calculated using Cronbach's alpha (Cortina, 1993). This is not to say that Cronbach's alpha is a measure of validity, merely a measure of how consistent the responses are to others in the same sample.

$$\alpha = \frac{K}{K-1} \left(1 - \frac{\sum s_j^2}{s^2} \right)$$

Where K is the number of questions, s^2 represents the variance of the observed total test scores, and s_j^2 is the variance of component j for the current sample.

CHAPTER 4

RESULTS

Descriptive Statistics

The sample study contains a total of 76 students. The ability groupings for the study group during the force unit have 26 high ability students (13 female, 13 male), and 25 for both the medium (12 female, 13 male) and low (16 female, 9 male) ability groupings. For the energy unit the ability grouping has 25 high ability (12 female, 13 male), 25 medium ability (12 female, 13 male), and 26 low ability students (17 female, 9 male).

Concept Tests

Figure 1 below is a histogram of student score distribution for the force concept inventory (FCI). The results of the exam show a cluster of students between 10 and 20 percent, while having a small population of high scoring test takers. According to Wells et al. (1995) the average high school test taker received an average of about 30% on the FCI. Hake (1998) measured the average high school score on the FCI to be 28% for 6,000 students. For the total sample of this study the average was 21.2% with 95% CI [19.6, 22.9]. The Cronbach's alpha for the FCI sample population of this study resulted in a reliability reading of 0.57, which equates to a poor reliability in the exam results. This poor reliability may be the result of student exhaustion or simply running out of time, both were common excuses of students leaving the class post exam.

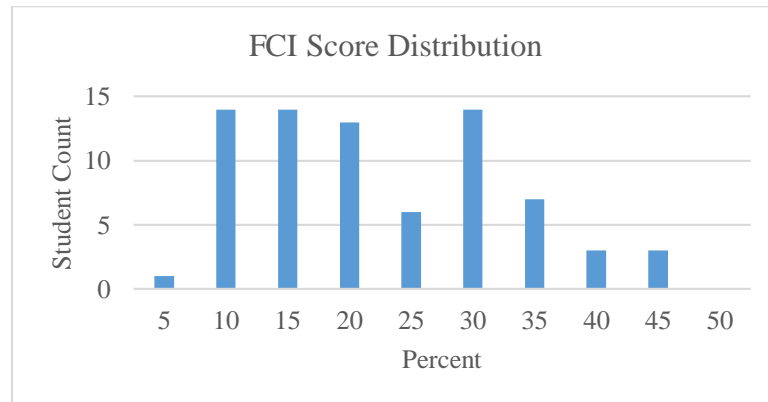


Figure 1. Student score distribution on the FCI.

The individual classes are compared in the box plot of Figure 2. There is not a significant effect between the class period and the FCI ($F(2,73)=1.297, p>0.1$) with a medium to small eta squared value of 0.034. Though it should be noted that the second class did seem to struggle on the FCI in comparison to the other classes.

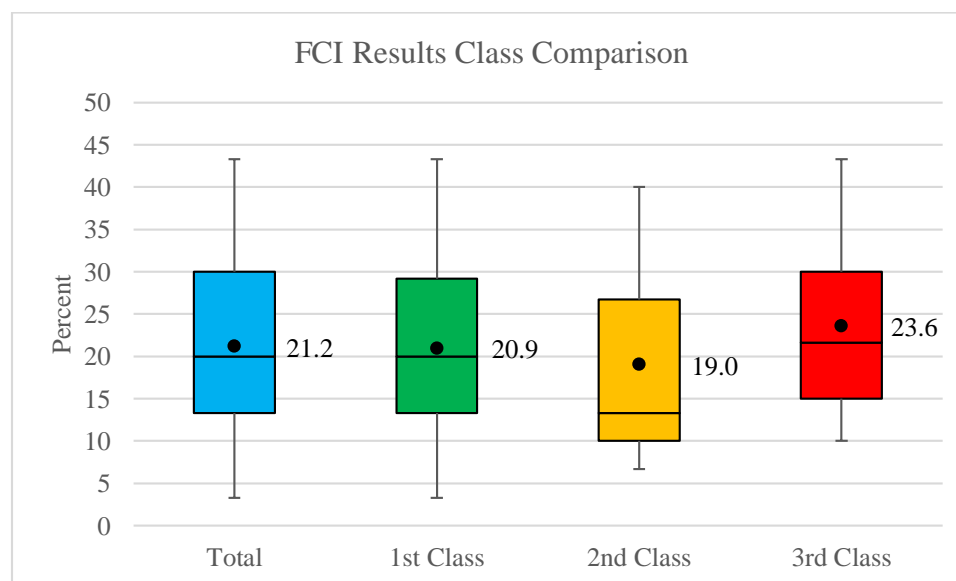


Figure 2. Percent score on FCI by class period.

The ability grouping, which was based on a combination of the FCI and current grade are plotted in a box plot next to the total sample score in Figure 3. The high and medium ability groupings share a similar upper bound for student on the FCI, yet the median and mean are higher for the high ability group. The low ability grouping has a lower 25% bound and median at the same overall score on the FCI. There is a significant effect of the ability groups average score on the FCI ($F(2,73)=30.8, p<0.001$) with very large eta squared value of 0.458. This is expected as the FCI was partially how the ability groups were constructed so there should be significance between these values.

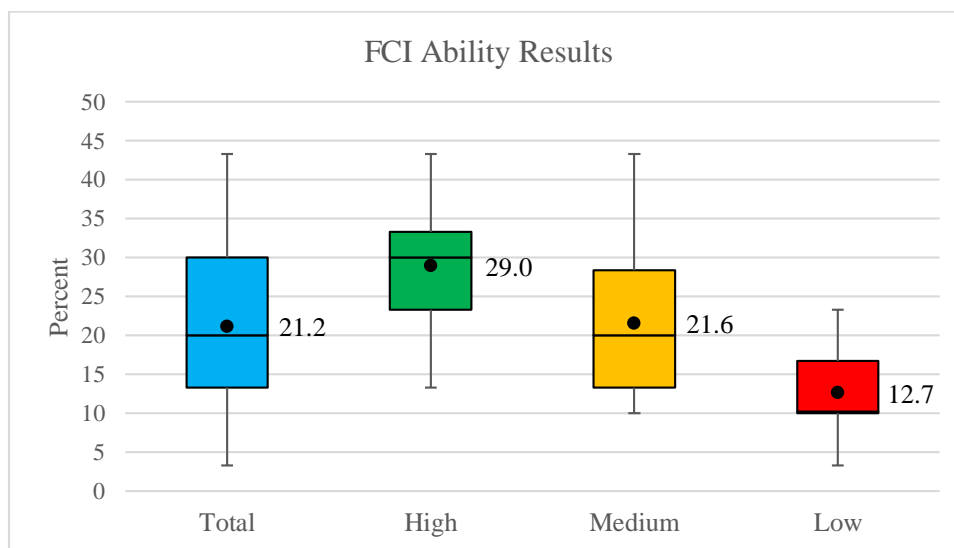


Figure 3. Percent score on FCI by ability group.

The energy and momentum conceptual survey (EMCS) measures student performance regarding energy and momentum as introduced by Singh and Rosengrant (2003). Their research of 186 students enrolled in an algebra based physics course resulted in an average of 24% with an Cronbach's alpha coefficient of 0.68 for the researched sample. This study participant's scores

resulted in an average score of 25.6% with 95% CI [23.9, 27.2]. The Cronbach's alpha coefficient was calculated to be 0.72 which is acceptable. The score distribution for the EMCS is shown in the histogram in Figure 4 below, compared to the FCI we notice a closer cluster of students between the 20 and 30 percent range.

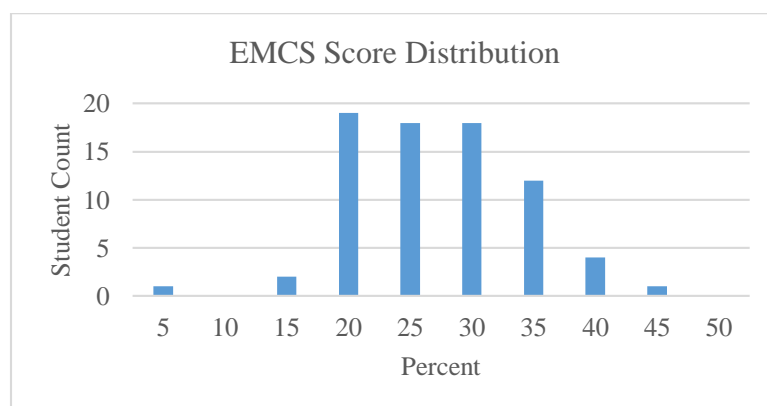


Figure 4. Student score distribution on the EMCS.

The EMCS scores by class are given in Figure 5. There is not a significant effect between the class period and the EMCS ($F(2,73)=1.175$, $p>0.1$) with a medium to small eta squared value of 0.031. This implies that each of the classes could be drawing from the same population.

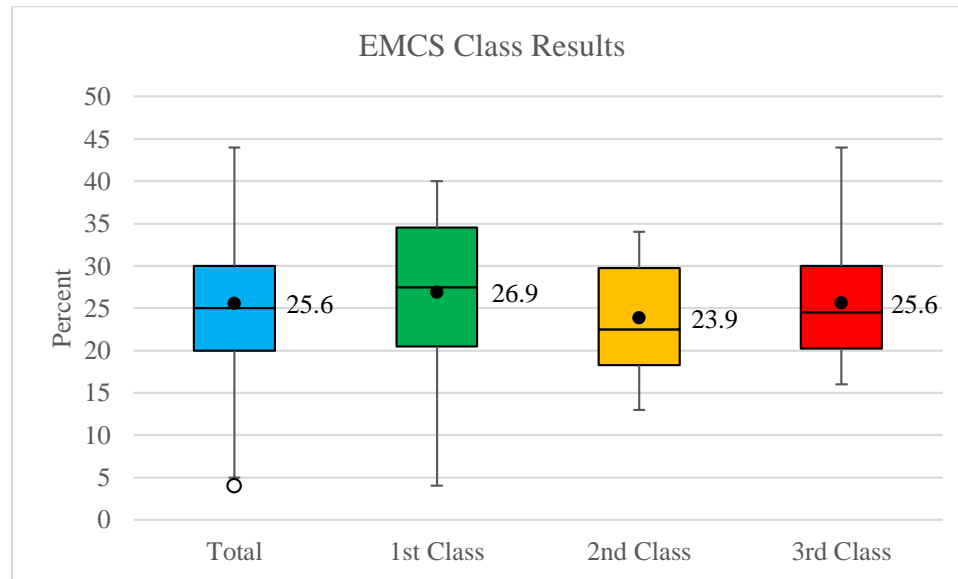


Figure 5. Percent score on EMCS by class period.

Box plots show the results for ability grouping on the EMCS in Figure 6. Less of a spread is noticed for the medium group on the EMCS when compared to the FCI. It should be noted that there are outliers at the extremes for the high and low ability groups, since this is the first instance of outliers, as previously defined. There is a significant effect of the ability groups average score on the EMCS ($F(2,73)=76.5$, $p<0.001$) with a very large eta squared value of 0.677. This significance is expected due to the new ability grouping construction.

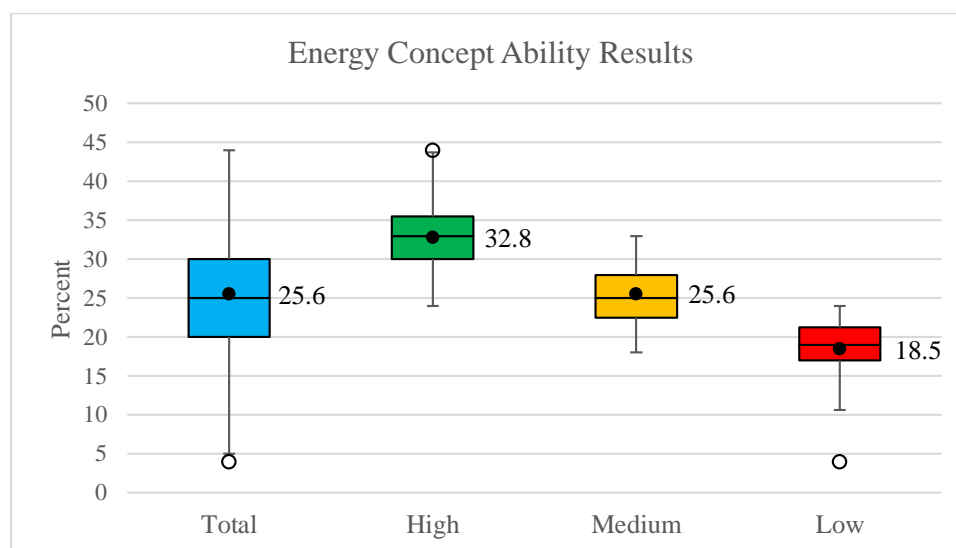


Figure 6. Percent score on EMCS by ability group.

Current Grade

The current grade was calculated based on in class exams, these exams are compared to the past class for simple comparison though the exams differ between questions, they do not differ in material covered. The pre-force current grade was calculated with a weighted average for the standards (and questions on the two summative kinematics exams). The results of the current pre-Force grade are given in Figure 7 for the current study and Figure 8 for the previous year's students. The Cronbach's alpha coefficient for the kinematics exams were calculated to be 0.80 and 0.81, which corresponds to good reliability for the exams, the previous year's exams resulted in acceptable Cronbach's alpha coefficients of 0.78 and 0.79. The distributions of the current students shows a peak value in the 3.5 range, while the previous year's students shows an almost exponential increase to the highest range. An unpaired t -test is performed between the years showing a significance ($t(120)=-3.88$, $p<0.001$) between means. The result is that the year of study saw a decrease in average score from the previous year of 0.51 with a 95% CI [-0.77,-

0.25] with a very large Cohen's d of -1.48. This implies that the previous year's students performed better on kinematics than the students in the current study, this is also shown in the histograms.



Figure 7. Pre-Force score distribution for current study.

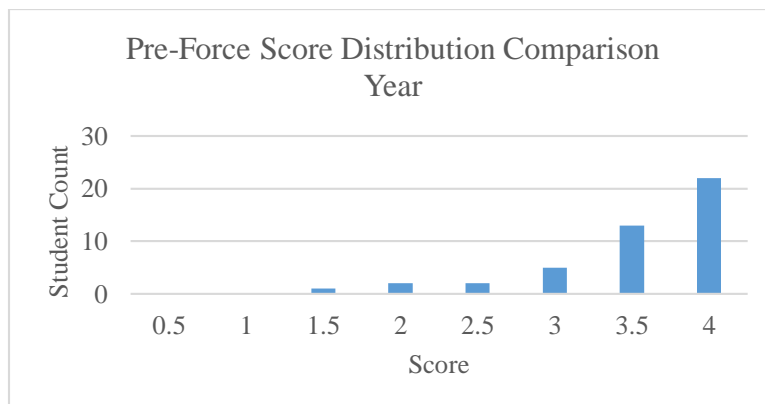


Figure 8. Pre-Force score distribution for last year's students.

The comparison between class periods pre-Force grade is given in Figure 9. There is not a significant effect between the class period and the pre-Force grade ($F(2,73)=0.746, p>0.1$) with a smaller eta squared value of 0.02 when compared to the concept tests. This again can be concluded as each class being statistically similar.

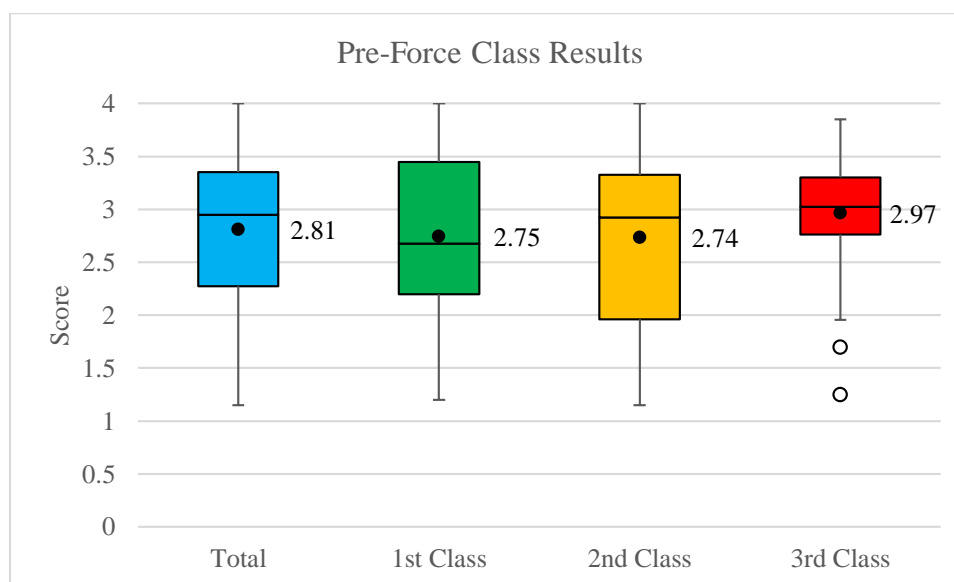


Figure 9. Score before Force unit by class period.

The current grade before the force unit is represented as box plots in Figure 10, which shows ability grouping with reference to last year's data. A clear difference is noticed between the ability results for the overall kinematic summative grades, however the low ability grouping does have a much larger spread than on the concept tests. In addition, an extremely statistical significant ($F(2,73)=41.8$, $p<0.001$) with very large eta squared of 0.53 was calculated for ability grouping and pre-Force grades. This is expected and mirrors the results on the FCI. Since ability groups were not constructed for the previous year, we can only compare the total results.

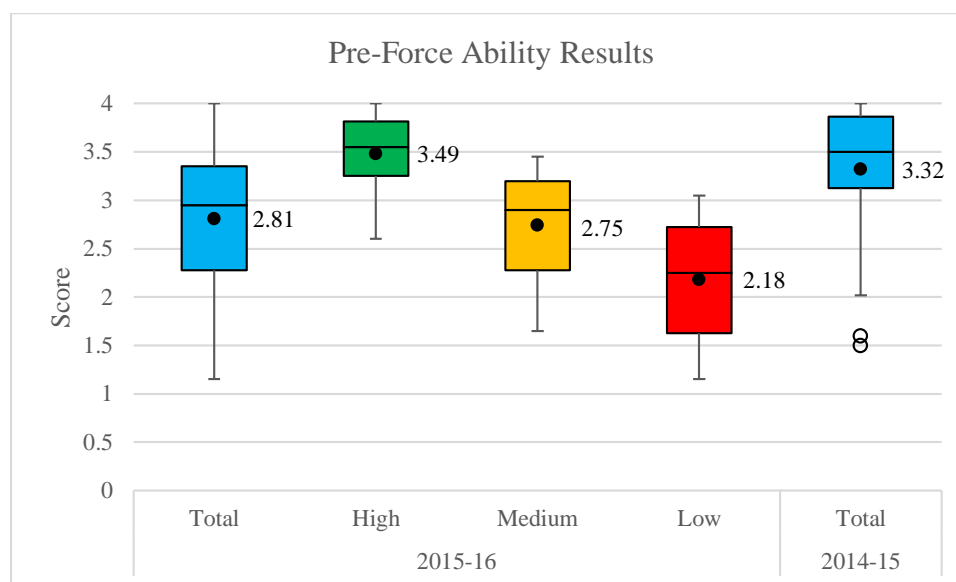


Figure 10. Score before Force unit by ability group and year.

Force Unit Survey Results

The force survey responses vary depending on the group or dyad configuration, which was taken for many different group interactions. The productivity responses refer to student survey response regarding how productive they felt their group was during class time. Figure 11 shows student response on productivity for group work in each of the configurations. Notice that the student choice group configurations have a lower percentage of strongly agree responses compared to the other configurations. We also notice a preference towards homogeneous grouping.

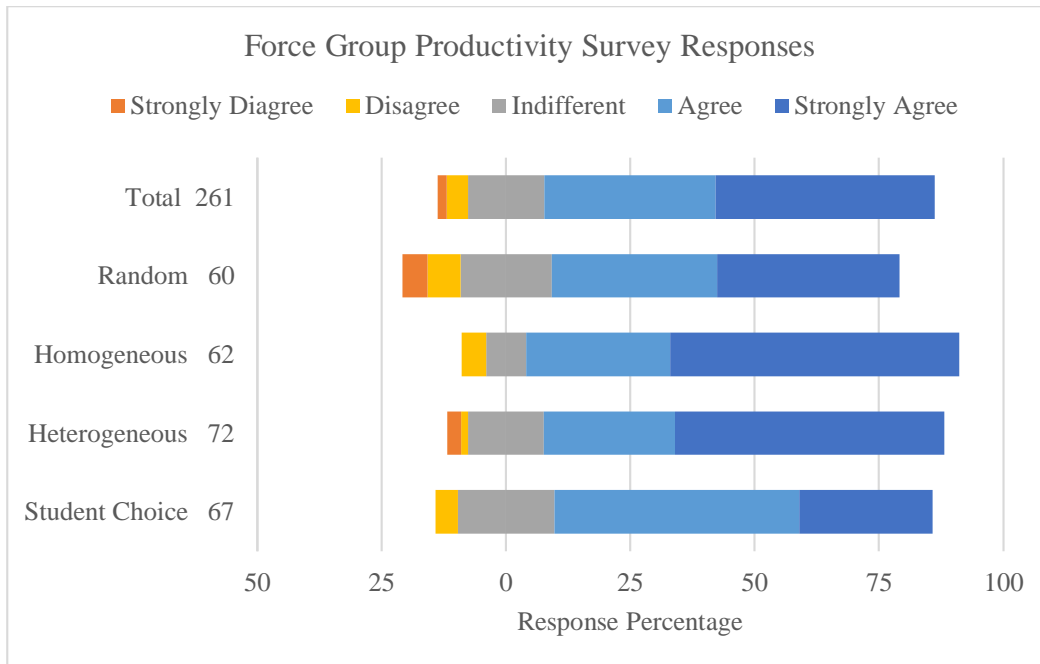


Figure 11. Survey response on group productivity by ability.

Student's responses regarding the benefit of the group to their understanding of the physics material is graphed in Figure 12. Similar to the productive measure, we see a student preference towards a homogeneous configuration and a bias against a high rating for the student choice configuration.

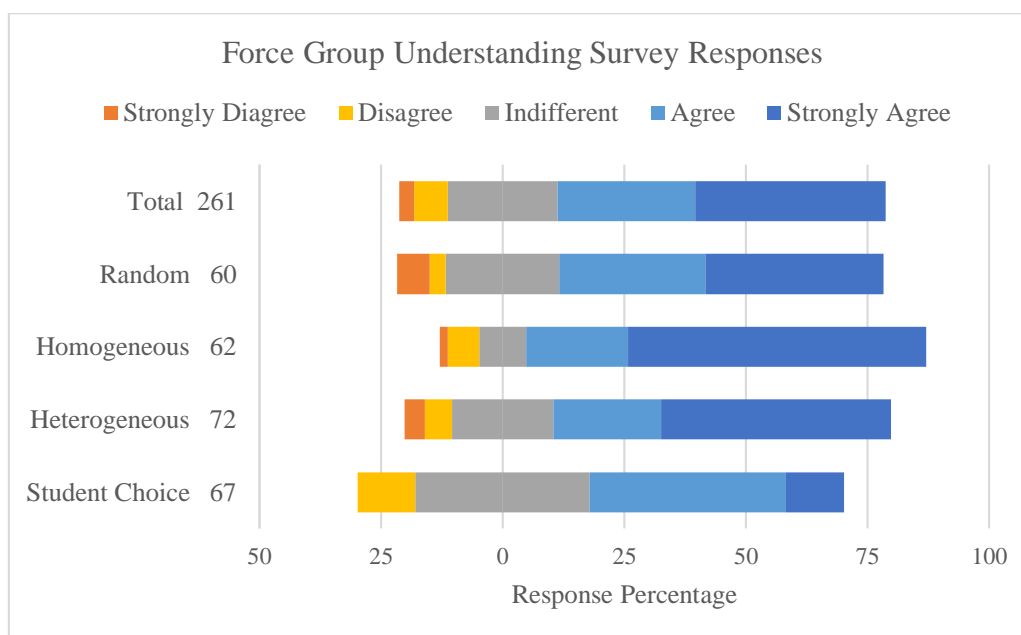


Figure 12. Survey response on group understanding by ability.

Finally, Figure 13 shows the student responses when asked if they would like to work with this group again. We notice an impressive 100% of the students decided that the student choice configuration is the best method. This student bias is exacerbated when looking at the Cronbach's alpha coefficient of the student responses regarding configurations, which resulted in lowest result for each group of 0.87, which implies that students were not reliable when responding for this group configuration. As a result of this uncertainty this type of group configuration was not studied for the energy unit. Heterogeneous configurations were overall the least preferred method, followed by homogeneous. The highest Cronbach's alpha was 0.95 for the random arrangement, followed by 0.94 for the heterogeneous and 0.91 for the homogeneous arrangements.

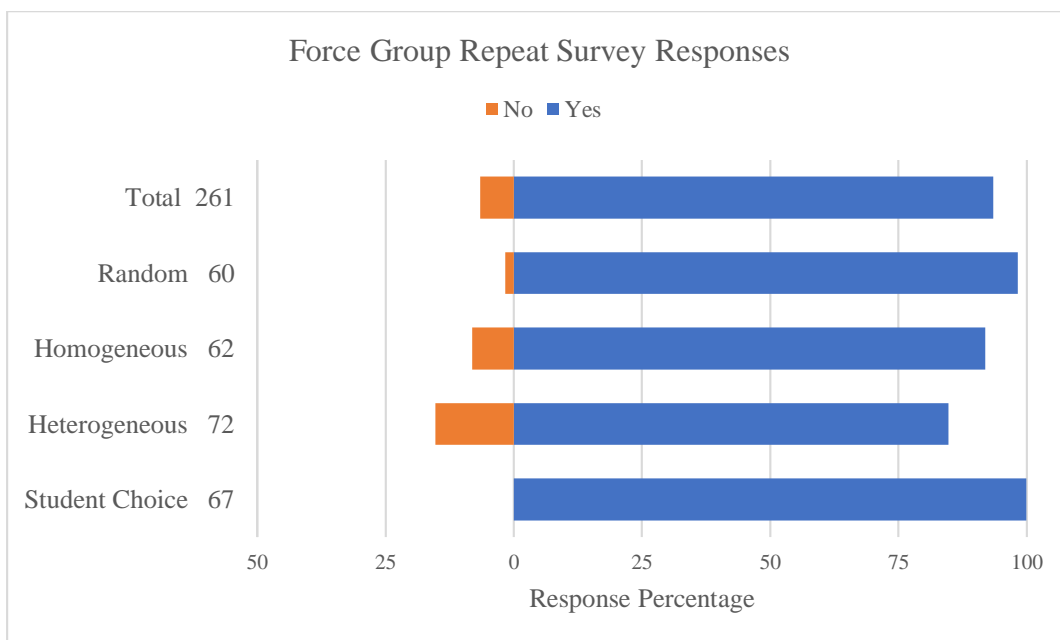


Figure 13. Survey response on group willingness to work together again by ability.

For dyad configurations, student survey responses regarding productivity resulted in a favorable measure for random and student choice, while the homogeneous configuration had the least favorable. This is displayed in Figure 14.

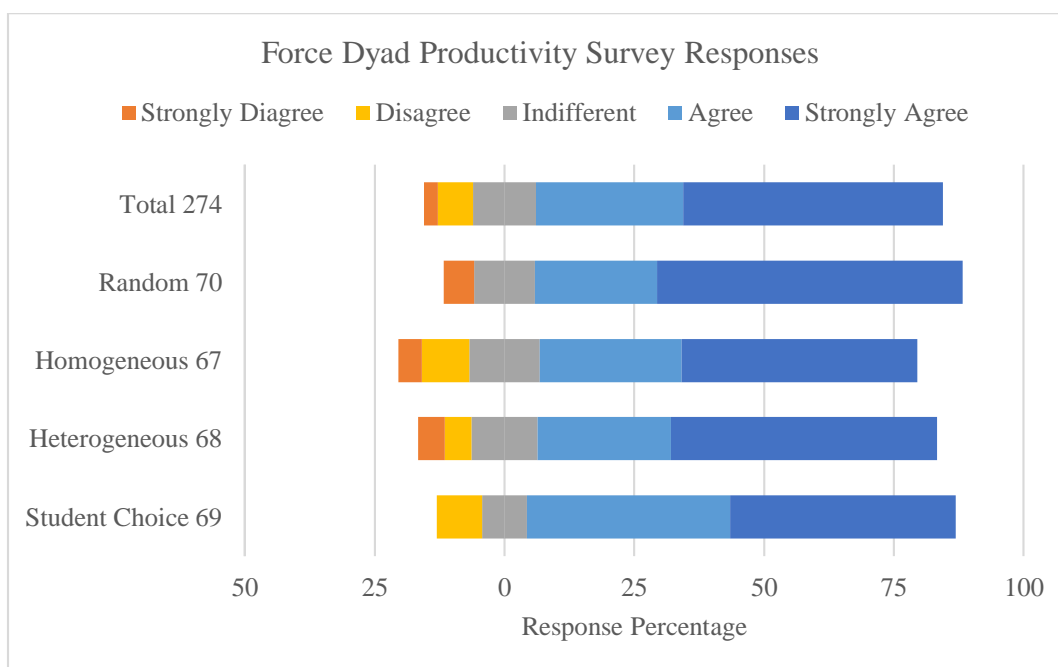


Figure 14. Survey response on dyad productivity by ability.

When asked to respond regarding how their partner influenced their understanding of the physics concepts, we notice a very favorable response rate for random configurations as seen in Figure 15. While no student responded with a strongly disagree response for the homogeneous configuration, it was still the least favored.

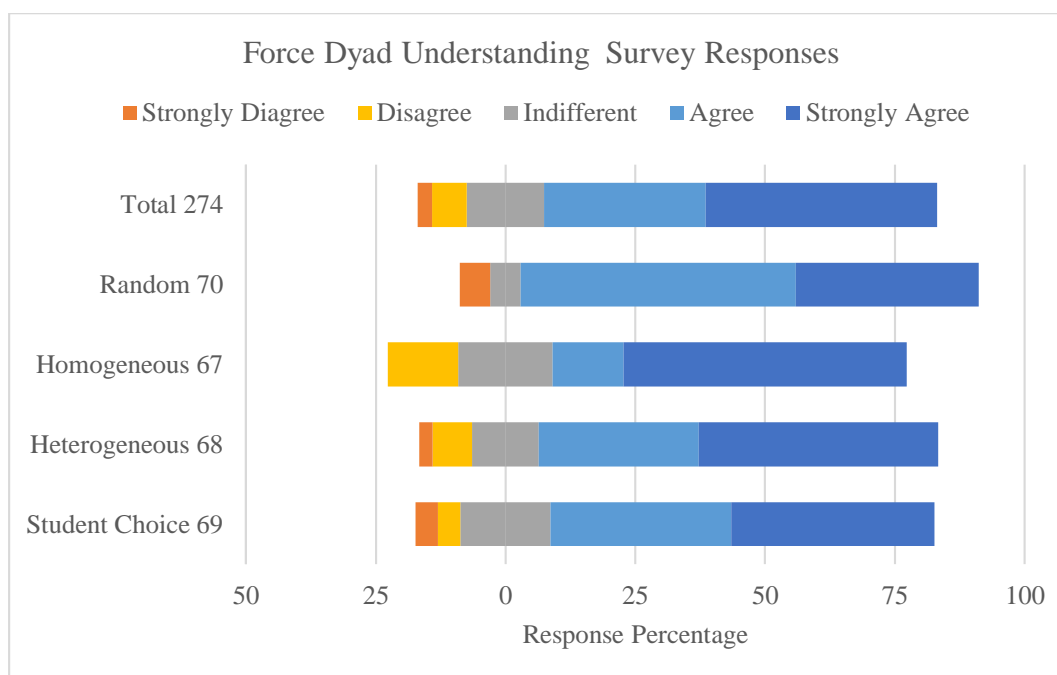


Figure 15. Survey response on dyad understanding by ability.

Finally, when asked to respond regarding willingness to work with their partner again, we notice an unfavorable rate for the homogeneous configuration followed by heterogeneous and student choice as seen in Figure 16. Remarkably the random partner configuration was the most favored in this measure. The Cronbach's alpha coefficient for all configurations were excellent, except for the student choice. The student choice Cronbach's alpha coefficient for student choice configuration was 0.76, which is an acceptable internal consistency. This lack of reliable results, in comparison to the other configurations, as well as the ambiguous relationship to our research questions resulted in eliminating student choice groups from the second portion of this study.

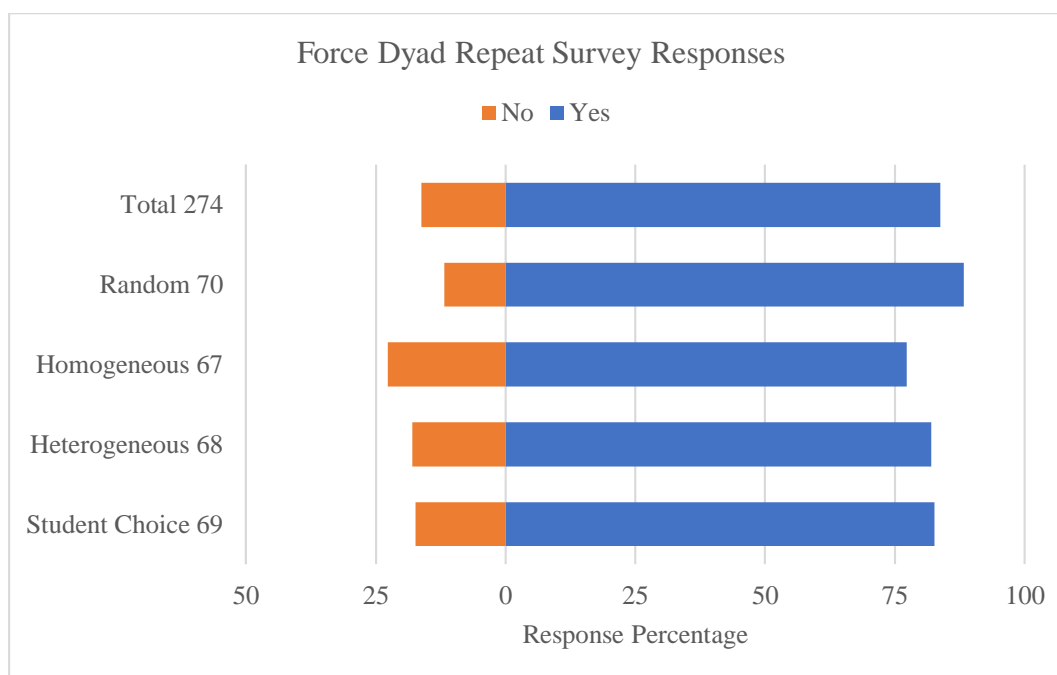


Figure 16. Survey response on dyad willingness to work together again by ability.

Results of Research Question

The research question looks to investigate the effect of different group arrangements with respect to the ability of the students. This question will be answered as a result of the two summative exams and survey responses.

Force Summative Exam

The force summative was given at the end of the unit and histogram of the results is given below in Figure 17, this can be compared to last year's class results diagramed in Figure 18. The Cronbach's alpha coefficient of the assessment was 0.84 for the current year and 0.83 for last year students, which indicates good reliability for both years. An unpaired t -test is performed

between the years resulted in no significance ($t(120)=-0.116, p>0.10$) between means. The result is that the year of study saw a decrease in average score from the previous year of 0.016 with a 95% CI [-0.29, 0.26] with a very small Cohen's d of -0.03. This implies that the current study population made gains on last year's students between the kinematics and force unit. This may be due to an increase in understanding by the students due to the group and dyad arrangement training, or have more time to work together in groups and dyads, or possibly better instruction.

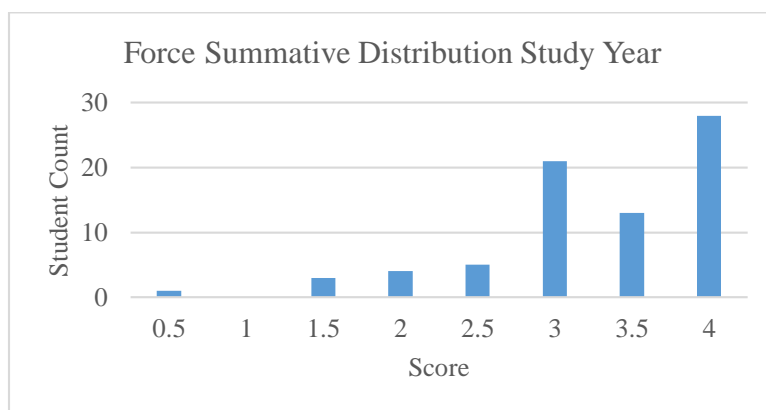


Figure 17. Force summative score distribution for the current study students.

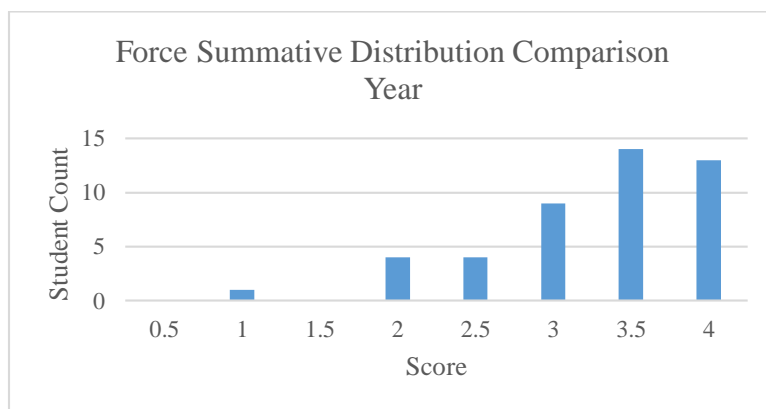


Figure 18. Force summative score distribution for last year's students.

The force summative grade is given in the box plot in Figure 19. There is not a significant effect between the class period and the Force summative score ($F(2,73)=1.225$, $p>0.1$) with a medium to small eta squared value of 0.032. This again can be intruded as each class being statistically similar.

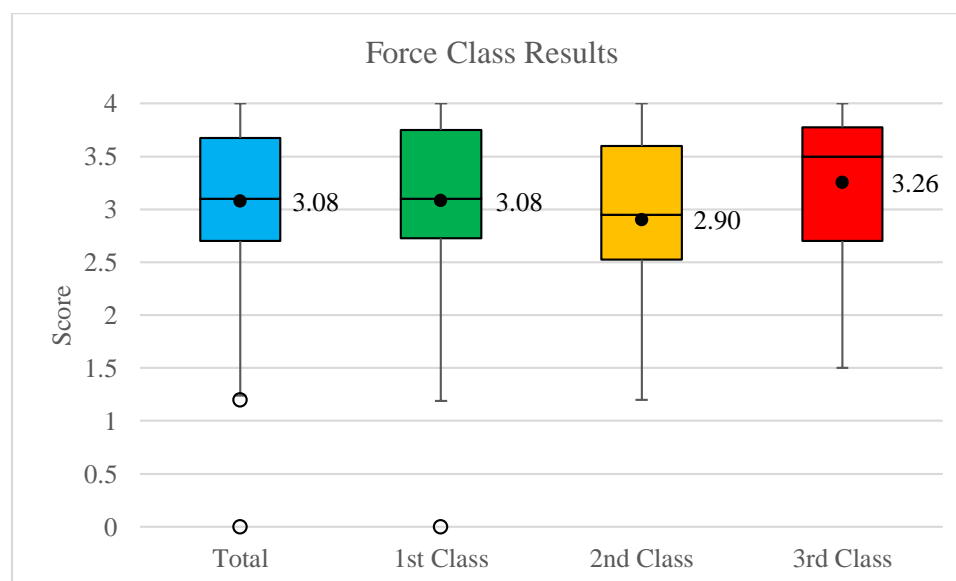


Figure 19. Score on Force summative by class period.

The force exam for the different ability groups compared to the total sample for our study as well as the previous year for comparison is given in Figure 20. A clear difference is noticed between the ability results for the overall Force summative grades, however the low ability grouping does have a much larger spread than on the concept tests, anywhere from a perfect score to zero. In addition, an extremely statistically significant ($F(2,73)=8.65$, $p<0.001$) with very large eta squared of 0.093 was calculated for ability grouping and force exam grade. So we

still have our previously defined ability groups, with the highest achieving still testing better than the medium and low ability groups.

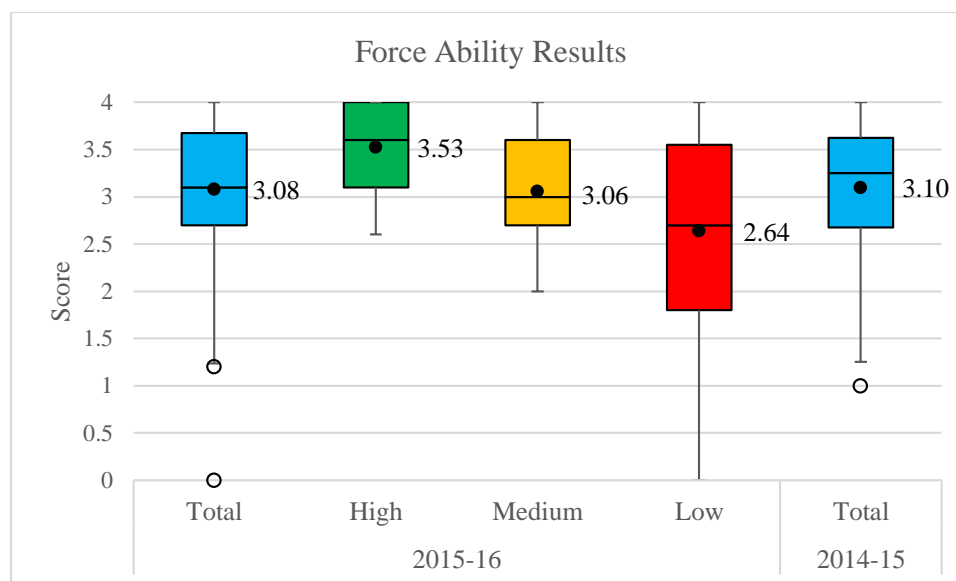


Figure 20. Score on the Force exam by ability group and year.

The average of the overall exam did show an improvement compared to the grade entering the exam. The paired *t*-test results when comparing the mean scores of the Pre-Force to Force summative exam scores are displayed in Table 6 for the total class and each of the ability groups. It should be noted that each ability group also saw an increase in their mean, however the low ability group saw the largest average normalized gain of 0.226 with medium effect size. The medium ability group also saw a significant increase in average normalized gain of 0.252 with very large effect size. While the high ability group did have a significant increase they still had an increase in the average normalized gain of 0.79 with medium effect size. The high ability group had the smallest chance to improve, yet still managed an increase in scores.

Table 6

Ability Grouping Paired Sample Test between Pre-Force and Force Scores

	Paired Mean Differences	95% CI of the Difference		<i>t</i> (Degrees of Freedom)	<i>p</i>	Cohen's <i>d</i>	Average Normalized Gains
		Lower	Upper				
Total 2015-16 (76)	0.268	0.128	0.408	3.809 (75)	$p < 0.001$	1.33	0.226
High (26)	0.040	-0.153	0.234	0.430 (25)	$p > 0.05$	0.04	0.079
Medium (25)	0.314	0.047	0.581	2.426 (24)	$p < 0.05$	3.07	0.252
Low (25)	0.458	0.191	0.725	3.536 (24)	$p < 0.01$	0.73	0.226
Total 2014-15 (46)	-0.226	-0.392	-0.060	2.735 (45)	$p < 0.01$	-1.37	-0.332

The previous year's students did not increase between the two exams, in fact we notice a statistically significant decrease in their average normalized gains of -0.332 with a very large effect size. This result points to an advantage in organized grouping and training of students.

Energy Summative Exam

For the energy unit, group configurations were held constant for the entire unit, yet each class had a different arrangement. The pre-energy current grade was calculated with a weighted average for the standards (and questions on the three summative exams). The results of the Energy summative are diagrammed in Figure 21 for the current study and Figure 22 for last year's students. Notice the similar shapes of both exam years, with the majority of student scoring well on the summative exam. An unpaired *t*-test is performed between the years showing no significance ($t(120) = -0.726$, $p > 0.1$) between means. The result is that the year of study saw a decrease in average score from the previous year of 0.070 with a 95% CI [-0.26, 0.12] with a small Cohen's *d* of -0.34. This implies that after the force exam that the current students have attained similar success for the course.

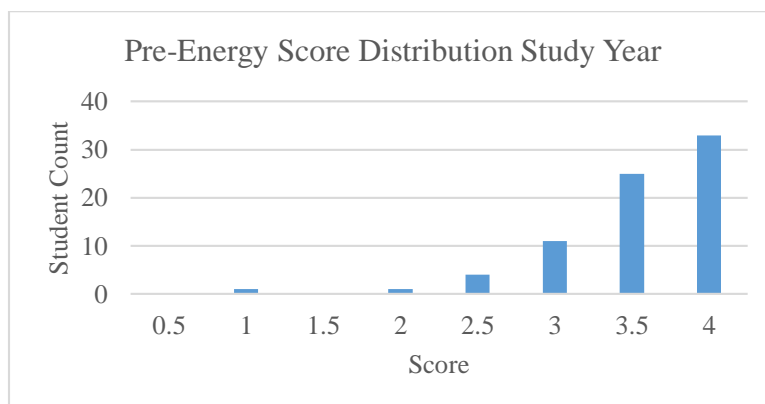


Figure 21. Pre-Energy score distribution for the current study students.

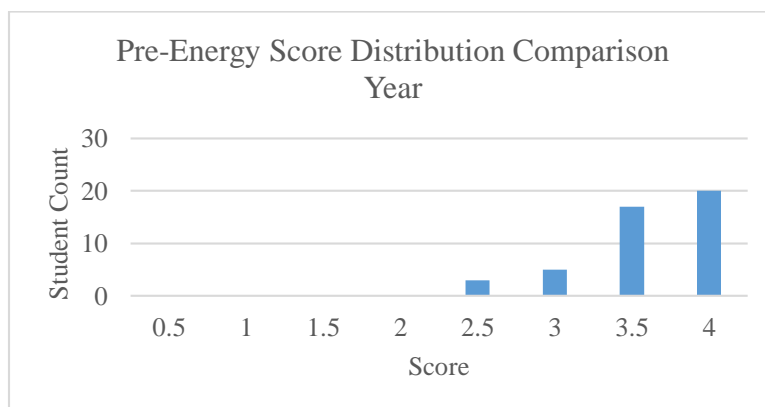


Figure 22. Pre-Energy score distribution for last year's students.

Figure 23 shows the scores before the Energy unit by class period. There is not a significant effect between the class period and the Pre-Energy grade ($F(2,73)=1.749, p>0.1$) with a medium eta squared value of 0.046. This again can be interpreted as each class being statistically similar.

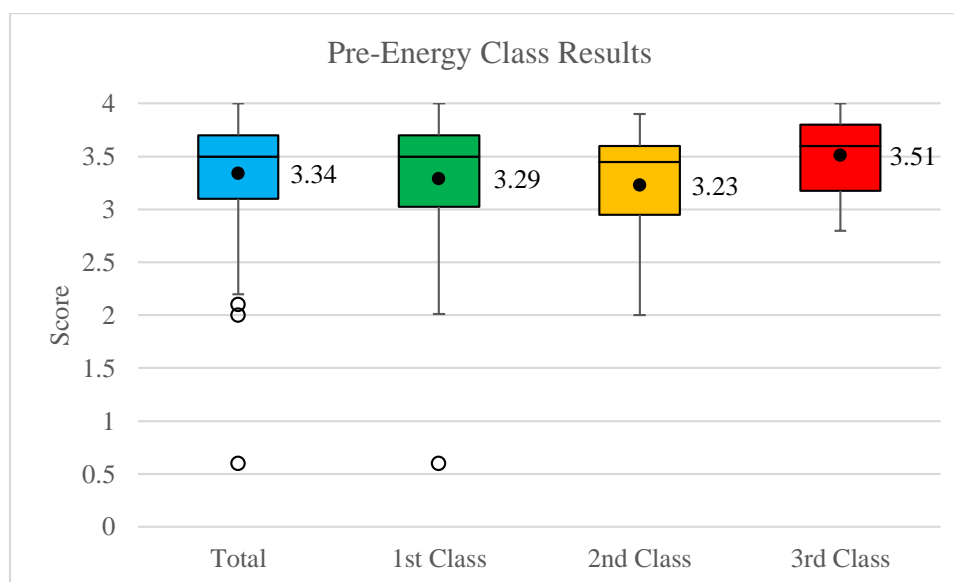


Figure 23. Score before Energy unit by class period.

The current grade before the energy unit is represented as box plots in Figure 24, which shows ability grouping with reference to last year's data. Though lower bound outliers exist for the total and medium ability grouping, the general trend between the ability groups is clearly demonstrated. The low ability group does have a wide range, stretching to the 75th percentile of the medium ability and into the 25th percentile of the high ability groupings. A clear difference is notice between the ability results for the overall grade. In addition, an extremely statistically significant ($F(2,73)=38.0, p<0.001$) with very large eta squared of 0.51 was calculated for ability grouping and pre-energy grades. This implies that the ability groups were effectively constructed.

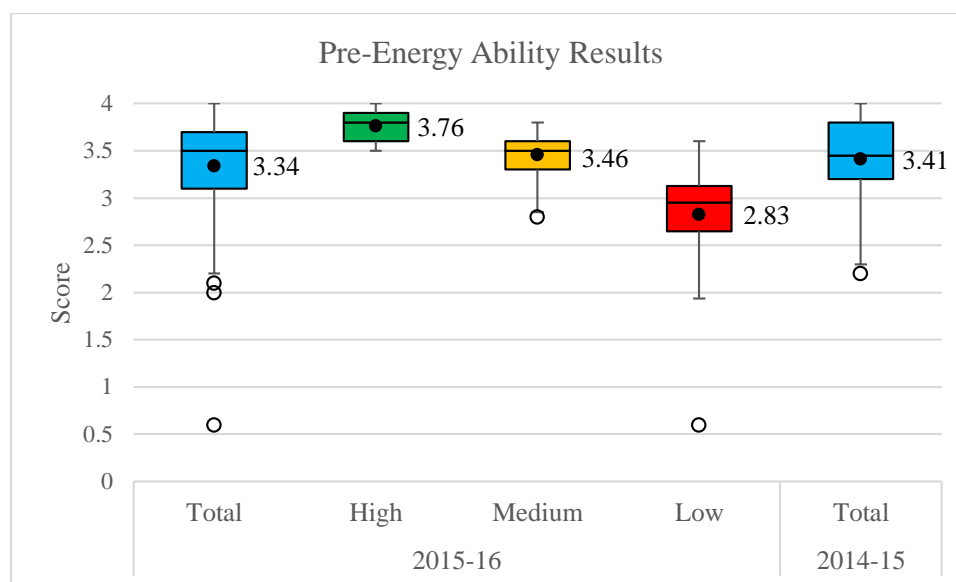


Figure 24. Score on before Energy unit by ability group and year.

The results of the Energy unit summative are depicted below in Figure 25 for the current study and Figure 26 for the previous year's students. The energy unit exam had an acceptable Cronbach's alpha coefficient of 0.71 for the current year, yet a poor 0.50 Cronbach's alpha coefficient for the previous year. An unpaired t -test is performed between the years resulted in no significance ($t(118.4)=-0.116, p>0.10$) between means. The result is that the year of study saw a decrease in average score from the previous year of 0.254 with a 95% CI [-0.33, 0.087] with a medium Cohen's d of -0.28. This implies that there is effectively no difference between them, so from the low performance during the kinematics unit to the current performance during the energy unit students in the study have effectively eliminated any advantage the previous year's students had.

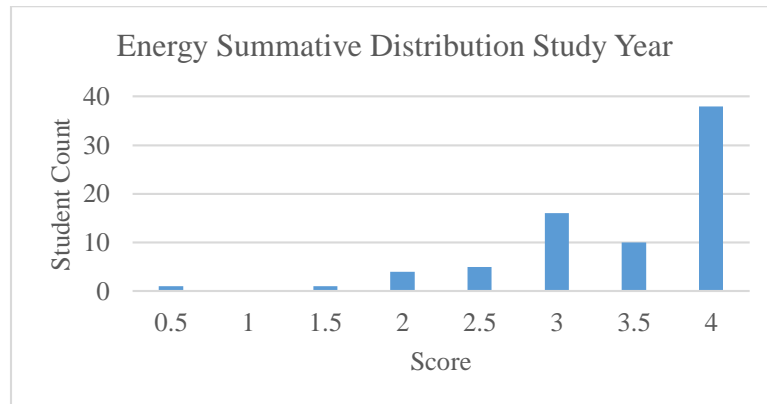


Figure 25. Energy summative score distribution for the current study students.

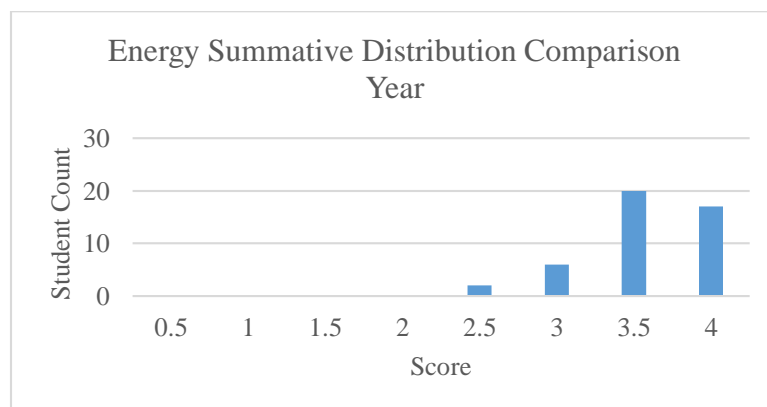


Figure 26. Energy summative score distribution for last year's students.

Figure 27 shows the Energy summative scores by class period. There is not a significant effect between the class period and the Energy summative score ($F(2,73)=0.021$, $p>0.1$) with a very small eta squared value of 0.001. This again can be interpreted as each class being statistically similar.

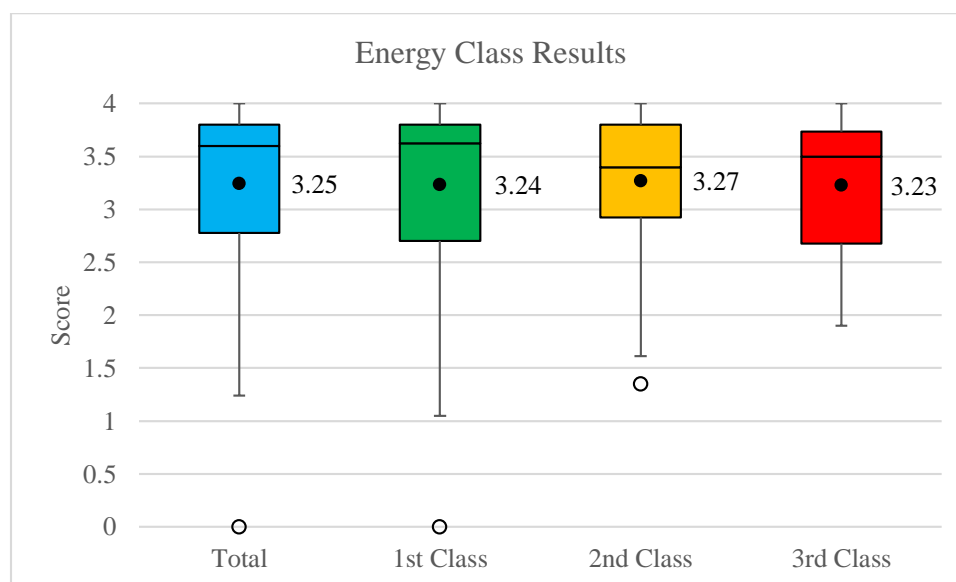


Figure 27. Score on the Energy summative by class period.

Figure 28 is a box plot of the ability subgroup results on the exam compared to the total sample. As was noted for the force exam, the low ability group resulted in the largest spread of data, yet the high ability showed low outliers and a 75th percentile that matched the 100th percentile at a perfect score. An extremely statistical significant ($F(2,73)=23.88$, $p<0.001$) with very large eta squared of 0.396 is noticed between the ability results for the overall Energy summative grade. Some statistically significant outliers occur for the high ability group, showing the struggles for that ability group to maintain their overall high level of performance.

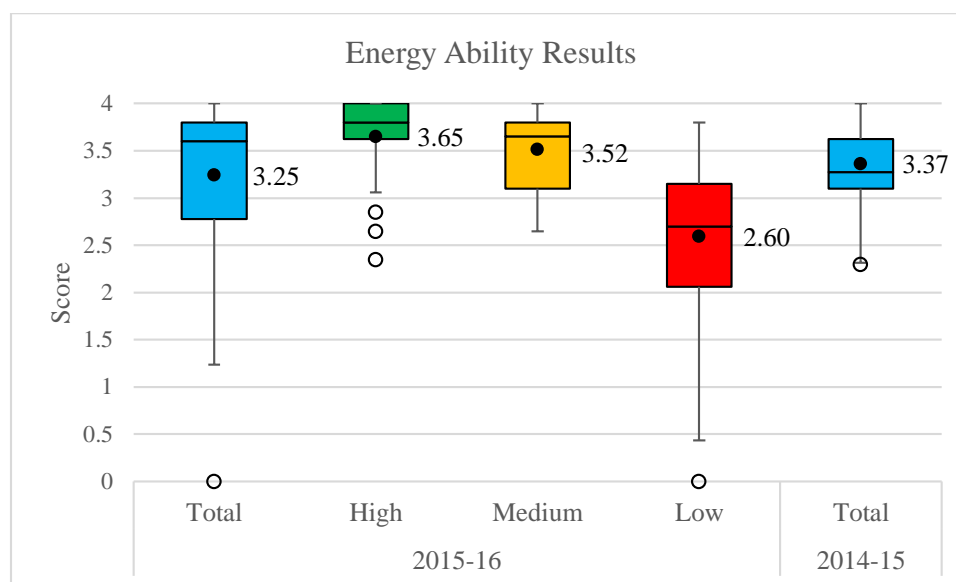


Figure 28. Score on the Energy exam by ability group and year.

The average of the overall exam did show an improvement compared to the grade entering the exam. The paired *t*-test results when comparing the mean scores of the pre-energy to energy scores are displayed in Table 7 for the total class and each of the ability groups based on their grouping configuration. Both years of students' scores decreased with no significance. Though the year of study decreased by double the amount as the previous year this maybe the result of the stronger class as discussed for the kinematics section. Statistical significance is measured in two instances, the heterogeneous grouping and the total low ability grouping. The total heterogeneous grouping saw a very significant decrease in the average normalized gains of -0.573 with a very large Cohen's *d*. The low ability grouping measured a significant decrease in the average normalized gain of -0.195, or a decrease in the mean of -0.229 with a medium Cohen's *d*. Gains between the exams were only measured for homogenous configuration with the high and medium ability groups, as well as the random configuration for the medium ability

grouping. Though no arrangement resulted in a net gain across each ability when looking at the average normalized gains. However, the homogeneous configuration appeared to be the most effective for all groups, and the heterogeneous configuration measured to be the least effective. The lack of success may be due to a class wide problem with the concepts of energy for the heterogeneous configuration especially.

Table 7

Ability Grouping Paired Sample Test between Pre-Energy and Energy Scores

		Paired Mean Differences	95% CI of the Difference		<i>t</i> (Degrees of Freedom)	<i>p</i>	Cohen's <i>d</i>	Average Normalized Gains
			Lower	Upper				
	Total 2015-16 (76)	-0.097	-0.201	0.008	-1.849 (75)	$p>0.05$	-0.290	-0.146
	Random (28)	-0.057	-0.267	0.153	-0.559 (27)	$p>0.1$	-0.124	-0.081
	Homogeneous (24)	0.040	-0.083	0.163	0.666 (23)	$p>0.1$	0.148	0.052
	Heterogeneous (24)	-0.279	-0.465	-0.093	-3.109 (23)	$p<0.01$	-1.064	-0.573
High Ability	Total (25)	-0.114	-0.284	0.056	-1.386 (24)	$p>0.10$	-0.905	-0.483
	Random (11)	-0.182	-0.551	0.187	-1.097 (10)	$p>0.10$	-0.910	-0.714
	Homogenous (6)	0.133	-0.025	0.291	2.169 (5)	$p>0.05$	19.19	0.444
	Heterogeneous (8)	-0.206	-0.415	0.003	-2.334 (7)	$p>0.05$	-2.429	-1.269
Medium Ability	Total (25)	0.058	-0.109	0.225	0.717 (24)	$p>0.10$	0.501	0.107
	Random (9)	0.183	-0.144	0.511	1.292 (8)	$p>0.10$	1.964	0.306
	Homogenous (6)	0.183	-0.212	0.578	1.193 (5)	$p>0.10$	2.082	0.333
	Heterogeneous (10)	-0.13	-0.389	0.129	-1.135 (9)	$p>0.10$	-0.780	-0.271
Low Ability	Total (26)	-0.229	-0.435	-0.022	-2.281 (25)	$p<0.05$	-0.590	-0.195
	Random (8)	-0.156	-0.671	0.359	-0.717 (7)	$p>0.10$	-0.207	-0.108
	Homogenous (12)	-0.079	-0.249	0.090	-1.028 (11)	$p>0.10$	-0.343	-0.071
	Heterogeneous (6)	-0.025	-1.263	0.013	-2.518 (5)	$p>0.05$	-4.563	-0.670
	Total 2014-15 (46)	-0.046	-0.154	0.061	-0.865 (45)	$p>0.10$	-0.599	-0.076

The large measured Cohen's *d* values can most likely be attributed to both small sample sizes as well as very small variances in student grades. Again the effect size is used as a

comparison internally for this study, so these large values merely show relative strength when compared to the other subgroups tested.

Though the material is not identical in each of the measures and the statistics are not particularly compelling, there is evidence to support that for this small sample that the most effectively way to positively influence all groups of ability is to use homogeneous grouping. These gains are highlighted by a positive trend in overall grade regardless of the ability grouping method for the entire sample when compared to previous year's classes. In addition, the statistically least effective method for grade improvement would be heterogeneous grouping. Again these results are not to be extended to the entire population of high school physics students without further testing.

Energy Unit Survey Results

The energy unit had groups divided into random grouping for the first class period, homogeneous grouping for the second class period, and heterogeneous grouping for the last class period. The general trends for each grouping resulted in similar Cronbach's alpha coefficients and responses on the survey for each question. For both group and dyad arrangements during the energy unit all Cronbach's alpha coefficients were considered excellent with the lowest being 0.90 and the highest being 0.95.

When looking at student preference regarding productivity for groups, we can investigate survey responses as shown in Figure 29. For the high ability grouping we notice a preference for the homogeneous configuration. For the medium ability grouping an advantaged is for the heterogeneous configuration, though it netted them the lowest gains academically. For the low ability grouping the homogeneous configuration is slightly preferred over random.

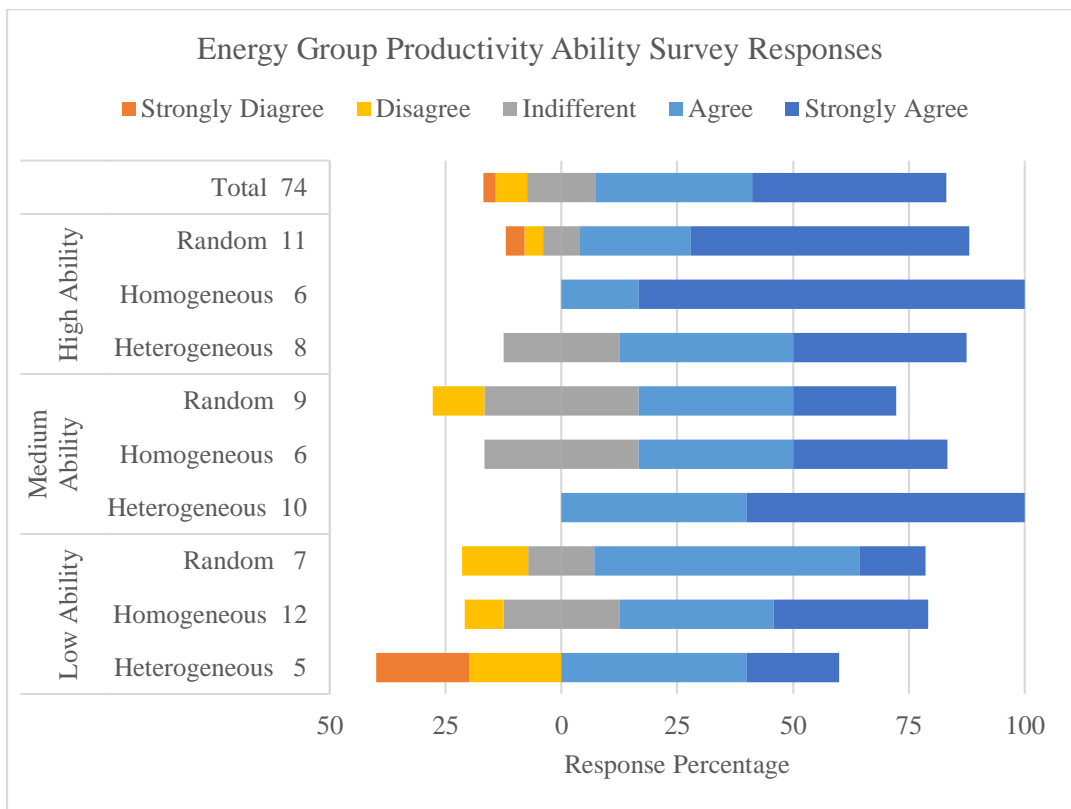


Figure 29. Survey response on group productivity by ability.

The student responses regarding their understanding are depicted in Figure 30. The understanding ratings for each ability grouping seems to match the response tendency for measure of productivity based on configuration. The exception being that the high ability group seems to prefer random more than heterogeneous configuration for second and third place after a clear preference towards homogeneous configuration. In addition the low ability group shows slight favor for the homogeneous configuration over the random when asked regarding understanding.

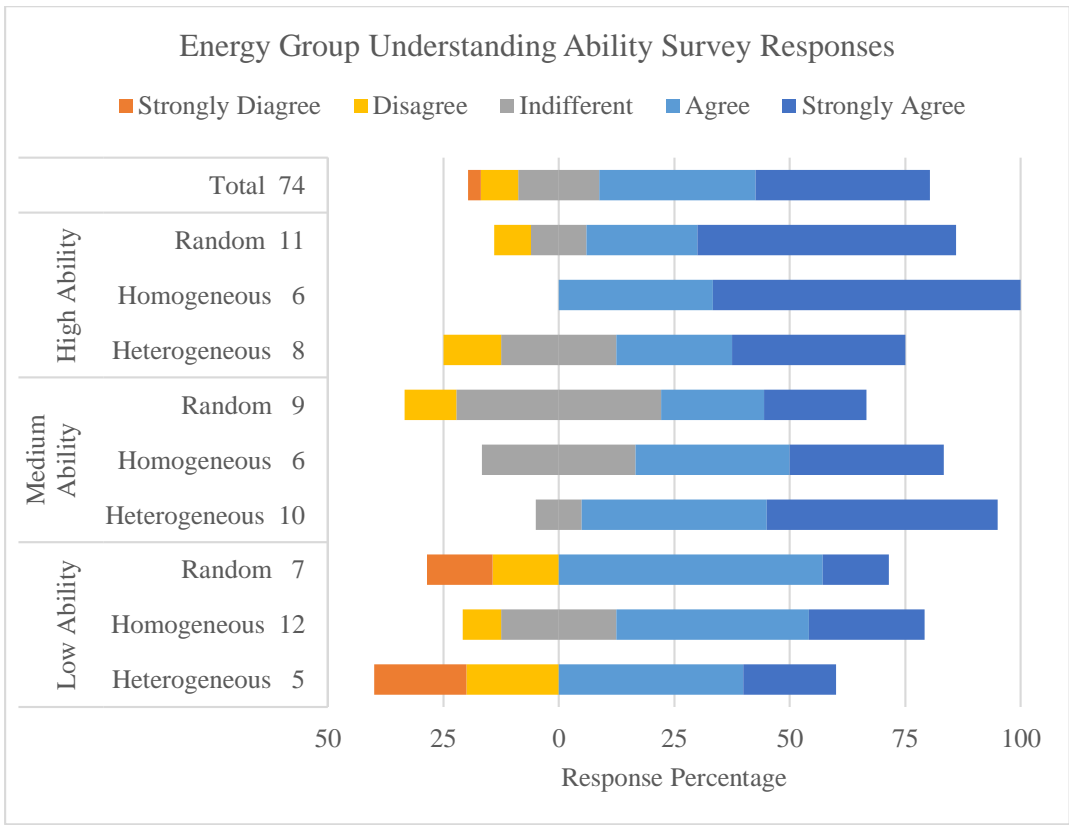


Figure 30. Survey response on group understanding by ability.

Finally, when asked if they would work in the grouping again the responses are shown in Figure 31. The high ability group appears to play no favorites, with a slight disdain for the random configuration. Similarly the medium ability group showed no preference between two configurations only marking down the homogeneous configuration. Finally, the low ability group had the most negative replies, especially for heterogeneous configuration.

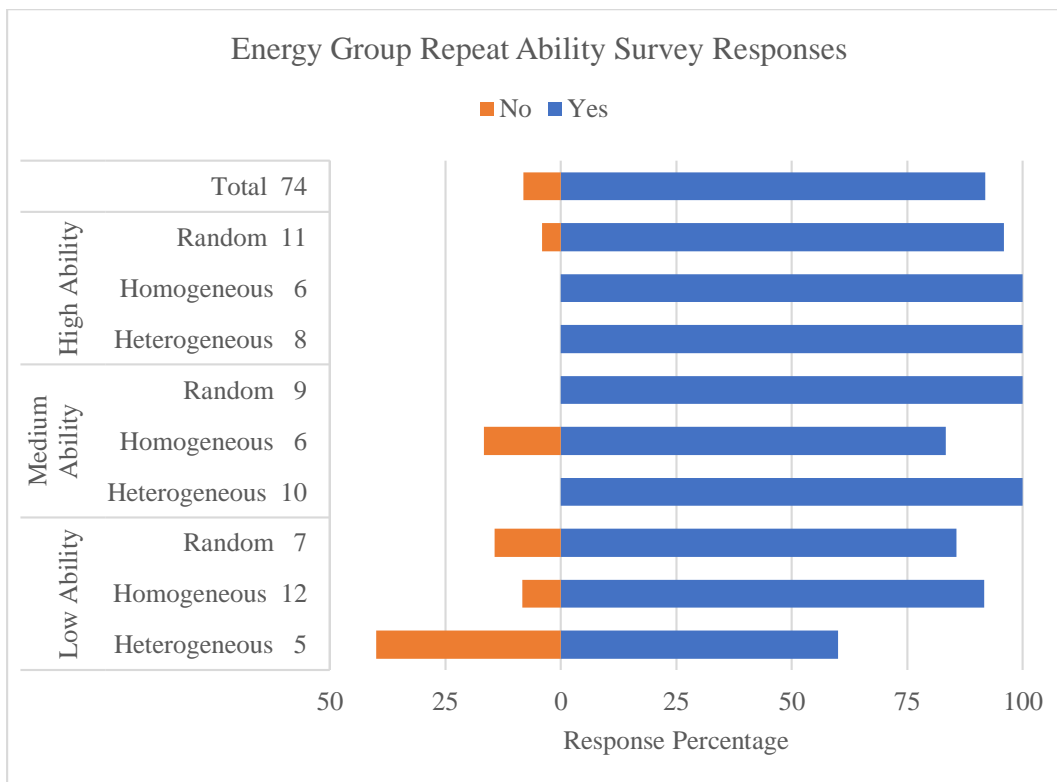


Figure 31. Survey response on group willingness to work together again by ability.

For dyad work the productivity survey responses based on ability grouping are shown in Figure 32. With the high ability group preferring the homogeneous configuration, while the medium ability group had no preference, and the low ability group preferred the heterogeneous configuration.

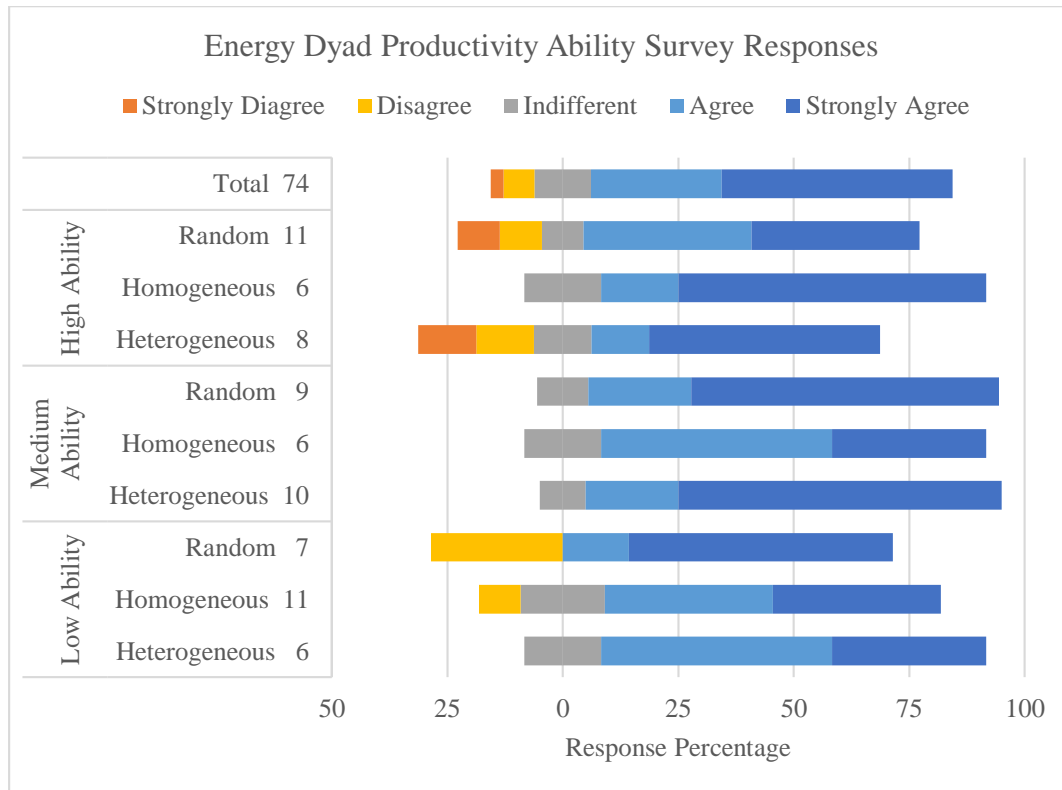


Figure 32. Survey response on dyad productivity by ability.

Student survey response are given in Figure 33 when asked regarding understanding. The high ability group still preferred the homogeneous configuration, while the medium and low ability groups preferred the heterogeneous configuration.

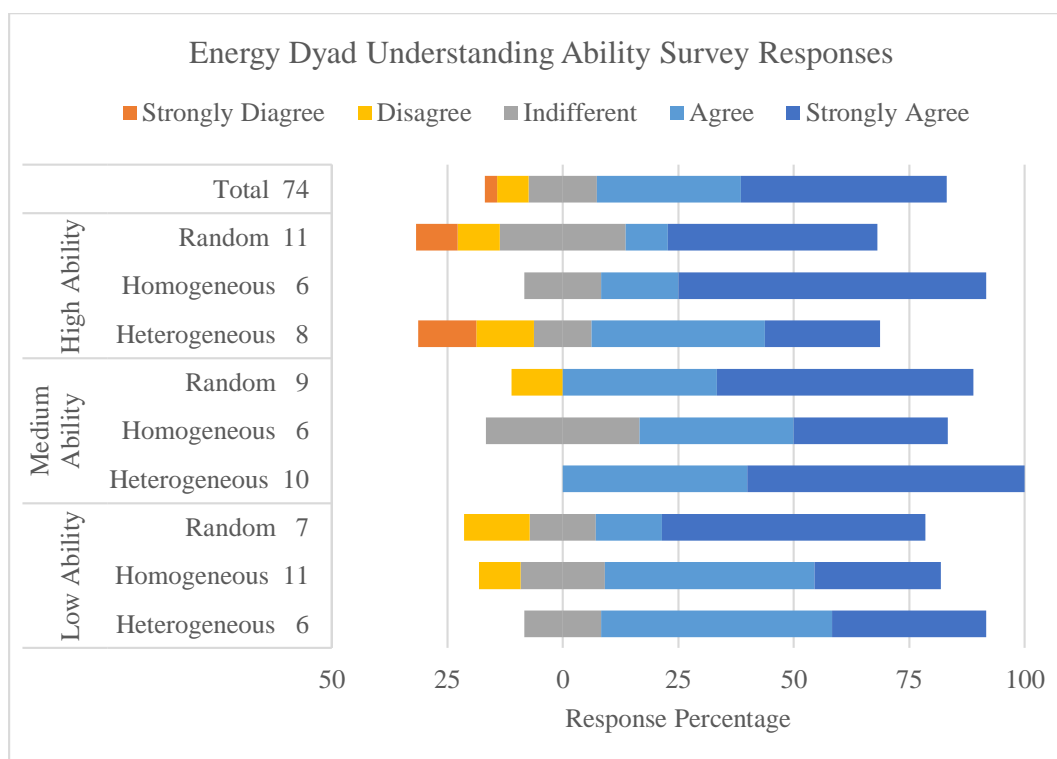


Figure 33. Survey response on dyad understanding by ability.

Finally, the survey responses when students were asked if they would like to work in this dyad again are shown in Figure 34. For the high ability group students preferred the random or homogeneous configurations, the medium ability student preferred heterogeneous, and the low ability students preferred either the homogeneous or heterogeneous configuration.

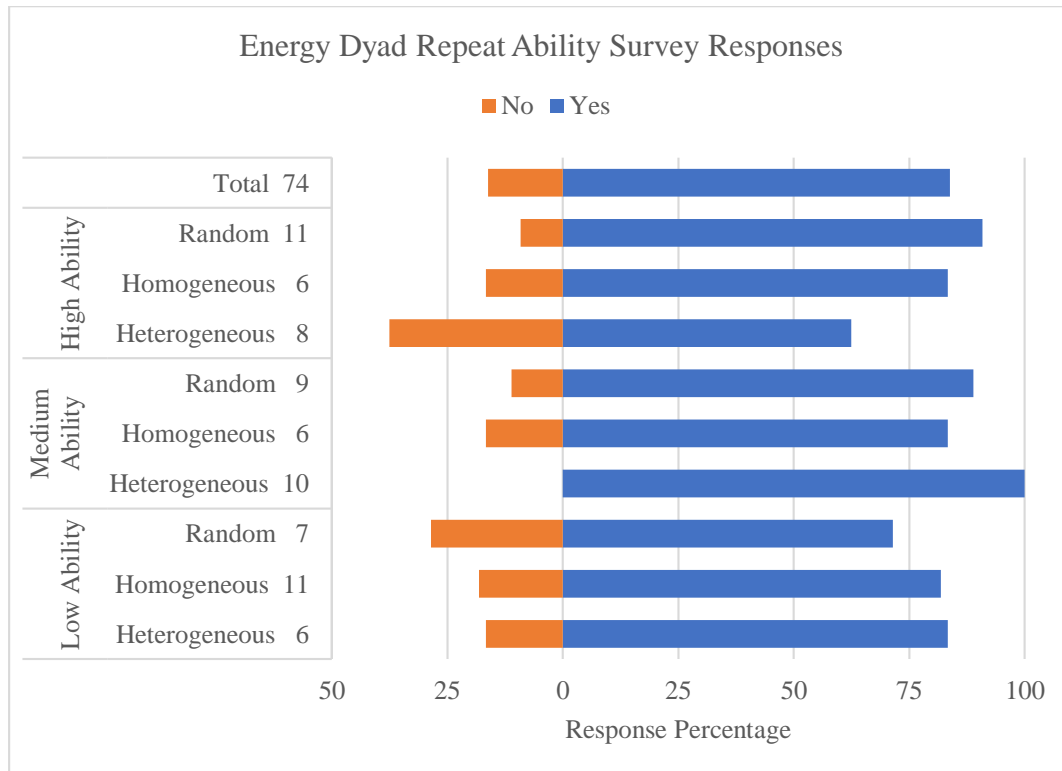


Figure 34. Survey response on dyad willingness to work together again by ability.

Based on the survey results, it appears that high and low ability students prefer to work in the homogeneous configuration, while medium ability students prefer heterogeneous configuration.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

Overview of Findings

This study investigated the composition of cooperative learning groups and dyads in the high school physics classroom. The study used statistical analysis of content exams as well as student survey results to determine the best composition of cooperative learning groups or dyads. The following conclusions based on the evidence presented for the high school honors physics class can be taken from this study:

1. There is evidence that homogeneous group and dyad arrangements result in the greatest academic achievement for all ability levels as shown in Table 7. This assertion is consistent with survey responses by students as shown in Figures 29 to 34.
2. There is evidence that heterogeneous group and dyad arrangements result in the lowest academic achievement for all ability levels as shown in Table 6.

These results are by no means overwhelming and require further study. It is probable that the low ability students may have preferred the homogeneous grouping due to an increase in instructor contact during the group and dyad sessions. In total the result of this study are consistent with recent findings in the physics classroom, Harlow et al. (2016) claimed that their results concluded that “instructors using team-based pedagogies should consider assigning the teams randomly, as it appears to be just as effective as sorted teams but requires significantly less

effort to implement.” While there is a possible slight advantage to homogeneous over random and heterogeneous grouping, these academic advantages may not outweigh the time commitment in group and dyad construction.

Limitations

Small sample size was a limitation for this study. With only 76 students participating in this study the small sample size was influenced by one instructor and this may have altered student behavior or survey responses. In addition, students may have succumb to the Hawthorne effect, where students may have subconsciously attempted to meet the intended outcome of the study (Bennett, 2015). With a large enough sample size these survey issues should be mitigated. According to a national survey of high school physics teachers conducted during the 2012-13 school year, the number of students enrolled in honors and regular physics is 746,000 (White & Tesfaye, 2014). With a confidence interval of 95% the margin of error for our sample of 76 students would be 11.24% when calculated with this 2012-13 population (Krejcie & Morgan, 1970). This is a lower bound assuming the population of honors and regular physics students has increased. For statistically significant results to be measured with a confidence interval of 95% and a margin of error of 3% the sample size during the 2012-13 school year would need to be 1066 students. Again this is a lower bound as the population most likely increased since the 2012-13 school year (Krejcie & Morgan, 1970).

The timeline for this study was too restrictive, group and dyad arrangements should be kept consistent for a longer length of time as noted by Zhang et al. (2017) for their study on college physics students. Students should be grouped in ability groups at the beginning of the year for their physics course and allowed to mature into a productive group.

Exceedingly small sample size educational studies with a relatively short study interval are problematic in educational research (Springer et al., 1999). However the small sample size does not discredit the usefulness of this data collection for implementation in my classroom to improve instruction. This experiment is not designed to inform policy but rather as a starting point for further study into the effectiveness of student arrangement by academic ability.

Finally, the medium ability group was often grouped in similar groups for both the heterogeneous and homogeneous methods. Thus this study cannot safely conclude any benefits or lack thereof for these methods for the medium ability group of students.

Recommendations

The following recommendations for further research are recommended based on the findings of this study:

1. Conduct a larger scale study that includes other types of physics classrooms from different schools across different districts or even countries. This will increase the chances of finding conclusive statistical results regarding ability performance.
2. Students should work with their groups and dyads for a full semester before any evaluations of the how the groups were constructed can be assessed with certainty.
3. Conduct a study with a more robust survey. Response questions should be aligned and guide students to one of the research questions. Increasing the number of Likert questions would increase the validity and allow for a more reliable conclusion. Ranking questions could be used when testing many

methods to show student preference. Possibly videotaping student interactions in groups could be implemented as a method to measure student performance while reducing student bias.

4. Investigate the effects of different student training methods when working in collaborative learning groups.
5. The materials students are engaged in during group and dyad work should also be studied. Are the materials appropriate for all types of groupings or do they work better for specific methods? Tiered questions with increasing difficulty could be used to measure group performance as a comparison for grouping methods.

The following recommendations for improving practice are recommended based on the findings of this study:

1. Student gain would be more accurately measured with a consistent exam.
Hence the FCI and EMCS should be given before and after the completion of the appropriate unit.
2. Four ability groups should be used instead of three so that medium ability students would work with different students when divided into heterogeneous and homogeneous groups or dyads.
3. Heterogeneous group construction should follow a method similar to the Swiss-system tournament. To implement this first organize the class in order of ascending ability. Then divide the class by the number of members of the group desired, so for this study it would be four sections. To form the groups simply take the top students in each section to form one group, and continue

down sections (Kujansuu, Lindberg, & Mäkinen, 1999). For example if there were 8 students and you wanted dyads, then 1 will be paired with 5, then 2 with 6, and so on. By implementing this method a better mix of students can be obtained.

This study provides insight into group and dyad configurations for the high school physics classroom. The recommendations are based on the findings of this study and are offered to improve instructor preparation for collaborative learning activities that best influence student academic achievement.

REFERENCES

- Amaria, R., Biran, L. A., & Leith, G. O. (1969, 02). Individual Versus Co-Operative Learning. *Educational Research, 11*(2), 95-103. doi:10.1080/0013188690110202
- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*(4), 369-406. doi:10.1037//0033-295x.89.4.369
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215. doi:10.1037//0033-295x.84.2.191
- Beebe, S. A., & Masterson, J. T. (2012). *Communicating in small groups: Principles and practices*. Boston: Allyn & Bacon.
- Bennett, T. (2015). Group work for the good: Unpacking the research behind one popular classroom strategy. *American Educator, 39*(10), 32-37. Retrieved from http://www.aft.org/sites/default/files/ae_spring2015bennett.pdf
- Bracey, G. W. (1994). Achievement in collaborative learning. *Phi Delta Kappan, 76*(3), 254.
- Burke, A. (2011). Group work: How to use groups effectively. *The Journal of Effective Teaching, 11*(2), 87-95.
- Carter, G., & Jones, M. G. (1994). Relationship between ability-paired interactions and the development of fifth graders' concepts of balance. *Journal of Research in Science Teaching, 31*(8), 847-856. doi:10.1002/tea.3660310807
- Chang, S. (2016, 04). LIGO detects gravitational waves. *Physics Today, 69*(4), 14-16. doi:10.1063/pt.3.3123

- Cohen, J. (2009). *Statistical power analysis for the behavioral sciences*. New York: Psychology Press Taylor & Francis Group.
- Cooper, J. (1990). Cooperative learning and college teaching: tips from the trenches. *Teaching Professor, 4*(5), 1-2.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98-104. doi:10.1037/0021-9010.78.1.98
- Coyle, V. C., Newman, D. L., & Connor, K. A. (n.d.). Innovative Instruction in STEM Education. *Handbook of Research on Active Learning and the Flipped Classroom Model in the Digital Age Advances in Educational Technologies and Instructional Design*, 309-332. doi:10.4018/978-1-4666-9680-8.ch016
- Crook, A. E., & Beier, M. E. (2010). When training with a partner is inferior to training alone: The importance of dyad type and interaction quality. *Journal of Experimental Psychology: Applied, 16*(4), 335-348. Retrieved from <https://search.proquest.com/docview/851223983?accountid=12846>
- Csernica, J., Hanyka, M., Hyde, D., Shooter, S., Toole, M., & Vigeant, M. (2002). *Practical guide to teamwork, version 1.1*. College of Engineering, Bucknell University.
- Davis, B. G. (2007). *Tools for teaching*. San Francisco: Jossey-Bass.
- Day, E. A., Arthur, W., & Shebilske, W. L. (1997, 11). Ability determinants of complex skill acquisition: Effects of training protocol. *Acta Psychologica, 97*(2), 145-165. doi:10.1016/s0001-6918(97)00019-x
- Devore, J. L. (2004). *Probability and statistics for engineering and the sciences*. Australia: Thomson-Brooks/Cole.

- Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C. (1996). The evolution of research on collaborative learning. In E. Spada & P. Reiman (Eds.), *Learning in humans and machine: Towards an interdisciplinary learning science*, 189–211. Oxford: Elsevier.
- Drew, D. E. (1996). *Aptitude revisited: Rethinking math and science education for America's next century*. Baltimore: The Johns Hopkins University Press.
- Elgort, I., Smith, A. G., & Toland, J. (2008, 02). Is wiki an effective platform for group course work? *Australasian Journal of Educational Technology*, 24(2). doi:10.14742/ajet.1222
- Felder, R.M., & Brent, R. (2001). Effective strategies for cooperative learning. *Journal of Cooperation & Collaboration in College Teaching*, 10(2), 69–75.
- Forman, E. (1989, 01). The role of peer interaction in the social construction of mathematical knowledge. *International Journal of Educational Research*, 13(1), 55-70.
doi:10.1016/0883-0355(89)90016-5
- Fuchs, L.S., Fuchs, D., Hamlett, C.L., & Karns, L. (1998). High-achieving students' interactions and performance on complex mathematical tasks as a function of homogeneous and heterogeneous pairings. *American Education Research Journal*, 35(2), 227–267.
doi:10.3102/00028312035002227
- Hake, R. R. (1998, 01). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64-74. doi:10.1119/1.18809
- Harlow, J. J., Harrison, D. M., & Meyertholen, A. (2016, 06). Effective student teams for collaborative learning in an introductory university physics course. *Physical Review Physics Education Research*, 12(1). doi:10.1103/physrevphyseducres.12.010138

- Heller, P., Keith, R., & Anderson, S. (1992, 07). Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving. *American Journal of Physics*, 60(7), 627-636. doi:10.1119/1.17117
- Heller, P., & Hollabaugh, M. (1992, 07). Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups. *American Journal of Physics*, 60(7), 637-644. doi:10.1119/1.17118
- Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The physics teacher*, 30(3), 141-158.
- Hooper, S., & Hannafin, M. J. (1988, 11). Cooperative CBI: The Effects of Heterogeneous versus Homogeneous Grouping on the Learning of Progressively Complex Concepts. *Journal of Educational Computing Research*, 4(4), 413-424. doi:10.2190/t26c-3fth-rnyp-tv30
- Hooper, S., Ward, T. J., Hannafin, M. J., & Clark, H. T. (1989). The effects of aptitude composition on achievement during small group learning. *Journal of Computer-Based Instruction*, 16, 102-109.
- Hooper, S. (1992, 01). Effects of Peer Interaction During Computer-Based Mathematics Instruction. *The Journal of Educational Research*, 85(3), 180-189. doi:10.1080/00220671.1992.9944435
- Illinois Report Card. (n.d.). Retrieved June 06, 2016, from <https://www.illinoisreportcard.com/>
- Jensen, J. L., & Lawson, A. (2011, 03). Effects of Collaborative Group Composition and Inquiry Instruction on Reasoning Gains and Achievement in Undergraduate Biology. *Cell Biology Education*, 10(1), 64-73. doi:10.1187/cbe.10-07-0089

- Johnson, D. W., Johnson, R. T., & Smith, K. A. (1991). *Cooperative Learning: Increasing College Faculty Instructional Productivity*. ASHE-FRIC Higher Education Report No.4. Washington, D.C.: School of Education and Human Development, George Washington University.
- Kirschner, F., Paas, F., & Kirschner, P. A. (2009, 03). Individual and group-based learning from complex cognitive tasks: Effects on retention and transfer efficiency. *Computers in Human Behavior*, 25(2), 306-314. doi:10.1016/j.chb.2008.12.008
- Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and psychological measurement*, 30(3), 607-610.
- Kujansuu, E., Lindberg, T., & Mäkinen, E. (1999). The stable roommates problem and chess tournament pairings. *Divulgaciones Matemáticas*, 7(1), 19-28.
- Kulik, J. A., & Kulik, C. C. (1987, 01). Effects of Ability Grouping on Student Achievement. *Equity & Excellence in Education*, 23(1-2), 22-30. doi:10.1080/1066568870230105
- Kulik, J. A., & Kulik, C. C. (1991). Research on ability grouping: Historical and contemporary perspectives. Storrs: University of Connecticut, National Research Center on the Gifted and Talented. (ERIC Document Reproduction Service No. ED 350 777)
- Lawrenz, F., & Munch, T. W. (1985, 03). Aptitude treatment effects of laboratory grouping method for students of differing reasoning ability. *Journal of Research in Science Teaching*, 22(3), 279-287. doi:10.1002/tea.3660220309
- Lawson, A. E. (1992, 09). Using reasoning ability as the basis for assigning laboratory partners in nonmajors biology. *Journal of Research in Science Teaching*, 29(7), 729-741. doi:10.1002/tea.3660290709

- Linchevski, L., & Kutscher, B. (1998, 11). Tell Me with Whom You're Learning, and I'll Tell You How Much You've Learned: Mixed-Ability versus Same-Ability Grouping in Mathematics. *Journal for Research in Mathematics Education*, 29(5), 533.
doi:10.2307/749732
- Lord, T. R. (1997, 03). A comparison between traditional and constructivist teaching in college biology. *Innovative Higher Education*, 21(3), 197-216. doi:10.1007/bf01243716
- Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & D'apollonia, S. (1996). Within-Class Grouping: A Meta-Analysis. *Review of Educational Research*, 66(4), 423.
doi:10.2307/1170650
- Mazur, E. (1996). *Peer Instruction: A User's Manual*. New York: Addison-Wesley.
- McDermott, C., Schaffer, P. S., & the Physics Education Group (2002). *Tutorials in Introductory Physics*. Englewood Cliffs: Prentice-Hall.
- Merali, Z. (2010, 03). Physics: The Large Human Collider. *Nature*, 464(7288), 482-484.
doi:10.1038/464482a
- Mugny, G., & Doise, W. (1978, 04). Socio-cognitive conflict and structure of individual and collective performances. *European Journal of Social Psychology*, 8(2), 181-192.
doi:10.1002/ejsp.2420080204
- Noddings, N. (1989, 05). Theoretical and Practical Concerns about Small Groups in Mathematics. *The Elementary School Journal*, 89(5), 607-623. doi:10.1086/461595
- O'Donnell, A. M., & King, A. (2009). *Cognitive perspectives on peer learning*. New York: Routledge.
- O'Donnell, A. M., Hmelo-Silver, C. E., & Erkens, G. (2006). *Collaborative learning, reasoning, and technology*. Mahwah, NJ: L. Erlbaum Associates.

- Opitz, M. F. (1998). *Flexible grouping in reading: Practical ways to help all students become stronger readers*. New York: Scholastic.
- Ormrod, J. E. (2000). *Educational psychology: Developing learners*. Upper Saddle River, NJ: Merrill.
- Physics Education Research and Development Group. (n.d.). Retrieved October 10, 2014, from <http://groups.physics.umn.edu/phyped/Research/CRP/on-lineArchive/ola.html>
- Rutherford, F. J., & Ahlgren, A. (1990). *Science for All Americans*. Oxford University Press.
- Shalaway, L. (2005). *Learning to teach: ... not just for beginners: The essential guide for all teachers*. New York: Scholastic.
- Shebilske, W. L., Regian, J. W., Arthur, W., & Jordan, J. A. (1992). A dyadic protocol for training complex skills. *Human Factors*, 34, 369 – 374.
- Simsek, A. (1992). The Impact of Cooperative Group Composition on Student Performance and Attitudes during Interactive Videodisc Instruction.
- Singh, C., & Rosengrant, D. (2003). Multiple-choice test of energy and momentum concepts. *American Journal of Physics*, 71(6), 607-617.
- Slavin, R. E. (1987). Ability Grouping and Student Achievement in Elementary Schools: A Best-Evidence Synthesis. *Review of Educational Research*, 57(3), 293. doi:10.2307/1170460
- Slavin, R. E. (2011). Instruction Based on Cooperative Learning. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of Research on Learning and Instruction*, 344-360. New York: Taylor & Francis.
- Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of Small-Group Learning on Undergraduates in Science, Mathematics, Engineering, and Technology: A Meta-Analysis. *Review of Educational Research*, 69(1), 21. doi:10.2307/1170643

- Tudge, J. (n.d.). Vygotsky, the zone of proximal development, and peer collaboration: Implications for classroom practice. *Vygotsky and Education*, 155-172.
doi:10.1017/cbo9781139173674.008
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA.: Harvard University Press.
- Watson, S. B., & Marshall, J. E. (1995a). Heterogeneous grouping as an element of cooperative learning in an elementary education science course. *School Science and Mathematics*, 95(8), 401-405. doi:10.1111/j.1949-8594.1995.tb10192.x
- Watson, S. B., & Marshall, J. E. (1995b). Effects of cooperative incentives and heterogeneous arrangement on achievement and interaction of cooperative learning groups in a college life science course. *Journal of Research in Science Teaching*, 32(3), 291-299.
doi:10.1002/tea.3660320308
- Webb, N. M. (1980, 01). A process-outcome analysis of learning in group and individual settings. *Educational Psychologist*, 15(2), 69-83. doi:10.1080/00461528009529217
- Webb, N. M. (1991, 11). Task-Related Verbal Interaction and Mathematics Learning in Small Groups. *Journal for Research in Mathematics Education*, 22(5), 366. doi:10.2307/749186
- Webb, N. M., Nemer, K. M., Chizhik, A. W., & Sugrue, B. (1998). Equity Issues in Collaborative Group Assessment: Group Composition and Performance. *American Educational Research Journal*, 35(4), 607. doi:10.2307/1163461
- Webb, N. M., Nemer, K. M., & Zuniga, S. (2002, 01). Short Circuits or Superconductors? Effects of Group Composition on High-Achieving Students' Science Assessment Performance. *American Educational Research Journal*, 39(4), 943-989.
doi:10.3102/00028312039004943

- Weld, K. (1999, 01). Perfect Problems And Homogeneous Groups Enhance Cooperative Learning In Abstract Algebra. *Primus*, 9(4), 355-364. doi:10.1080/10511979908965941
- Wells, M., Hestenes, D., & Swackhamer, G. (1995, 07). A modeling method for high school physics instruction. *American Journal of Physics*, 63(7), 606-619. doi:10.1119/1.17849
- White, S., & Tesfaye, C. L. (2014). High School Physics Courses & Enrollments: Results from the 2012-13 Nationwide Survey of High School Physics Teachers. *Focus On. Statistical Research Center of the American Institute of Physics*.
- Woolfolk, A. E. (2007). *Educational psychology*. Boston, MA: Pearson A and B.
- Zhang, P., Ding, L., & Mazur, E. (2017, 01). Peer Instruction in introductory physics: A method to bring about positive changes in students' attitudes and beliefs. *Physical Review Physics Education Research*, 13(1). doi:10.1103/physrevphyseduces.13.010104

APPENDIX A

FORCE UNIT SUMMATIVE EXAM

Honors Physics Quiz 3: Standard B (Newton's Laws)

B1 _____ B2 _____ B3 _____

Name: _____ Period: _____ Date: _____

Show all work. Draw a box around your final answer for all non-multiple choice questions.

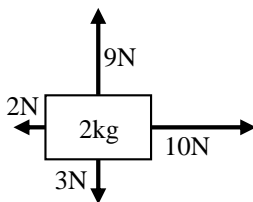
1. Three forces act on an object. If the object is in equilibrium, which of the following must be true: (B1)

- I. The vector sum of the three forces must equal zero.
 - II. The magnitudes of the three forces must be equal.
 - III. All three forces must be parallel.
- a. I only
 - b. II only
 - c. I and III only
 - d. II and III only
 - e. I, II, and III

2. A 5kg severed zombie head slides down a hill with a constant speed of 25m/s for 5s. What is the net force acting on the now dead undead head? (B1)

- a. 49 N
- b. 125N
- c. 0 N
- d. 25 N
- e. Not enough information

3. For the free body diagram find the acceleration in the x and y directions. (B2):



4. A block is being pushed across the room by Lolo in order to block King Egger from attacking him. The block has a weight of 50 N and Lolo pushed it with a force of 80 N on a frictionless surface. (B2)

a) Draw and label the forces on the block with a free body diagram:

b) Find the acceleration of the block.

5. The engine of an 800 kg pumpkin car produces a forward thrust of 1510 N. The forces of air resistance and friction combine to make an opposing force of 316 N. (B2)

a) Draw and label the forces on the pumpkin car with a free body diagram.

b) At what rate will the pumpkin car accelerate?

6. A presidential candidate is pushing on a 50 kg ballot box. When starting from rest he manages to reach a speed of 1.3m/s in just 4 seconds. What force did the gubernatorial candidate push the box assuming a frictionless surface? (B2)

7. A maximum force of 260 N is applied to a 125 kg coffin before it moves from rest on the level cemetery ground. (B1,B3)

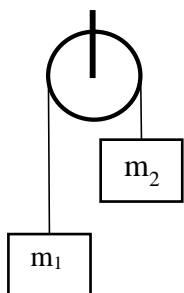
- a) What type of friction is this demonstrating? _____
- b) What is the coefficient of friction for this situation?

8. A 4.5kg block is sliding across the table at an initial speed of 4 m/s, yet is brought to rest by a frictional force. The friction coefficient is 0.5 between the block and the table. (B2,B3)

- a) What type of friction is this demonstrating? _____
- b) What is the frictional force acting on the block?

c) How far will the block slide before coming to rest?

9. An Atwood device is drawn below, for each situation find the acceleration of the blocks.
(B1,B3)



a) $m_1=5\text{kg}$, $m_2=5\text{kg}$

b) $m_1=2\text{kg}$, $m_2=5\text{kg}$

10. Draw and label the action reaction force pairs for the following situations. (B3)

a) Bowling ball pushes bowling pin forward.

b) Enclosed air particles push balloon wall outwards.

c) Stan pulls on a rope for a tug-o-war against Keith. (Hint: This situation has more than one set of action reaction pairs)

APPENDIX B

ENERGY UNIT SUMMATIVE EXAM

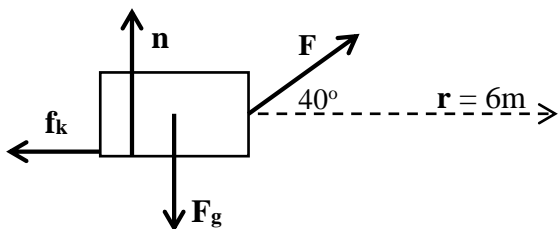
Honors Physics Quiz 4: Standard C (Work, Energy, and Power)

C1_____ C2_____ C3_____ C4a_____ C4b_____ C5_____

Name: _____ Period: _____ Date: _____

Show all work. Draw a box around your final answer for all non-multiple choice questions.

1. Find the value of work done for each force drawn below on the 20kg box and discuss what the sign indicates (C1)



a) $F = 100\text{N}$:

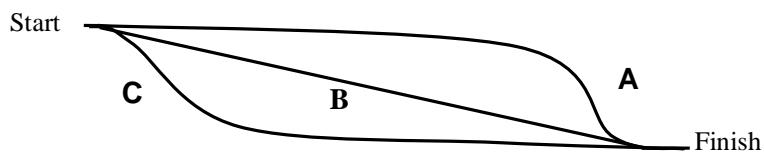
b) $f_k = 50\text{N}$:

c) F_g :

d) n :

For 2 and 3 use the following:

Three marbles are released from rest and roll down the friction free inclines labeled below:



2. Which path results in the marble obtaining the highest speed? (C4a)

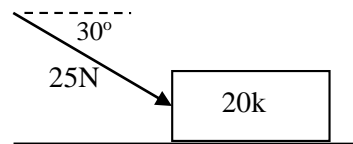
- A
- B
- C
- All paths have the same final speed.

3. Which path results in the shortest time for the marble to finish? (C5)

- A
- B
- C
- All paths take the same time.

4. A 20 kg box is pushed with a force of 25N at an angle of 30° below the horizontal for a distance of 15 meters.

a) How much work is required to push the box? (C1)



b) What is the power if it takes 110 seconds to move the box? (C5)

5. A block pushed at a constant speed of 2.5 m/s by an applied force of 30 N requires how much power? (C5)

6. A 5.0 kg ball is held over the edge of the John Hancock Center. If the ball had 16,856 J of energy, what is the height of the building? (C3)

7. The engine of a 1.5×10^6 kilogram train does 2×10^6 J of work to move the train from rest on a level surface. What speed will the train reach after the work is done? (C1,C2)

8. A 0.50-kilogram pendulum bob is released from some initial height such that the speed of the bob at the bottom of the swing is 0.60 m/s. What is the initial height of the bob? (C4a)

9. A 160 meters tall rollercoaster is has carts moving at a speed of 0.9 m/s at its highest point. What speed are the carts traveling at a height of 40 meters? (C4a)

10. A 0.1 kg ball at rest on the top of a 1.5 meter tall track. At the base of the track the ball reaches a speed of 4.9 m/s.

a) How much energy is lost? (C4b)

b) Where did this energy go? (C4a)

11. A 3.5 kg bowling ball is accidentally thrown in the air with a velocity of 3.2 m/s from a height of 0.80 meters. If the ball lost 20 joules of energy to air resistance what speed will the ball hit the bowling lane? (C4b)

APPENDIX C

CONTEXT-RICH PRACTICE PROBLEM EXAMPLE

External Forces and Conservation of EnergyNames: _____

Solve the following question with your partner. You may need your equation/conversion sheet and calculator (watch the mode).

You are driving your car along a straight level road. Suddenly, you see a van with what appears to be an adolescent werewolf surfing on top run a red light and enter the intersection just ahead of you. You slam on your brakes and skid in a straight line to a stop, leaving skid marks 100 feet long. A policeman observes the whole incident and gives a ticket to the van for running a red light (among other violations). He also gives you a ticket for exceeding the speed limit of 45 mph. When you get home, you read your physics book and estimate that the coefficient of kinetic friction between your tires and the road was 0.60, and the coefficient of static friction was 0.80. You look in your owner's manual and find that your car weighs 2,050 lbs. Will you fight the traffic ticket in court? Prove your case below.

Bonus: What would change about your case if instead of a level road you were driving up a hill that made an angle of about 10° with the horizontal? Use the back of the sheet to prove your case.