The Effects of Developing Kinematics Concepts Graphically Prior to Introducing Algebraic Problem Solving Techniques

James Archambault, Theresa Burch, Michael Crofton, Angela McClure

Principal Investigator: Dr. Robert Culbertson
Arizona State University

Action Research required for the Master of Natural Science degree
with concentration in physics.
July 2008

Table of Contents

Abstract …………………………………………………………………………. 2

Rationale ………………………………………………………………………… 3

Literature Review ……………………………………………………………….. 3

Method …………………………………………………………………………….. 5

Procedure for Treatment ………………………………………………………... 6

Data Analysis …………………………………………………………………… 8

Timeline …………………………………………………………………………… 9

Results (treatment group in general physics) ……………………………………… 9

Conclusion ……………………………………………………………………… 22

Investigator 1 Field Trial Report ………………………………………………….. 23

Investigator 2 Field Trial Report ………………………………………………….. 29

Investigator 3 Field Trial Report ………………………………………………….. 42

Investigator 4 Field Trial Report ………………………………………………….. 55

Implications for Further Research ……………………………………………… 66

Acknowledgements ……………………………………………………………... 67

Works Cited ……………………………………………………………………….. 68

Appendices ……………………………………………………………………… 69

Abstract

Kinematics concepts in physics have been modeled algebraically, graphically and with motion maps. However, the investigators have noticed that despite the many representations available to them, students show a strong preference for solving kinematic problems algebraically. Students were required to solve kinematics problems exclusively utilizing graphical representations prior to being introduced to methods of algebraic solutions. Results indicate that students who received the treatment chose graphical methods for solving kinematics problems more frequently, and demonstrated stronger gains in understanding kinematics concepts overall than those students who did not receive the treatment.

The effects of developing kinematics concepts graphically prior to introducing algebraic problem solving techniques

Principal Investigator: Robert Culbertson
Co-Investigators: Jim Archambault, Theresa Burch-Lococo, Michael Crofton, and Angela McClure

Rationale

Kinematics, the study of motion, is amongst the first topics introduced in most introductory physics classrooms throughout this country because the major ideas in these units, position, velocity and acceleration, are incorporated into almost all other topics in physics. Without a solid understanding of these important concepts, students lack the foundation necessary to succeed as physical concepts become more abstract and more complicated to model mathematically. Because of the fundamental nature of kinematics in the whole of physics, any improvement in student understanding of these concepts creates the possibility of an improved understanding of almost all of the rest of the physics concepts that students will encounter throughout the year.

This study attempted to improve conceptual understanding of kinematics concepts by having students solve problems with graphical representations prior to acquiring the ability to solve similar problems algebraically. By spatially exploring these problems before abstracting the concepts to an algebraic level, we emphasized the conceptual understanding with the students. We hoped that this method of teaching kinematics would be successful in creating a deeper understanding of the concepts of kinematics than the modeling program has done. We examined our results in part with the TUG-K and the FCI that are common assessment tools in physics education (Jackendoff, 1999, and Hestenes, 1992).

Literature Review

Students who receive a good grade in an introductory physics course have demonstrated their ability to algebraically find the answer to a multitude of physics problems, but many of these same students, upon closer inspection, are lacking a proper conceptual understanding of even the most fundamental physics concepts (McDermott, 1993). Further, students who do not examine their preconceived notions of the physical world, confront them with evidence, and change their way of thinking may be able to complete a mathematical algorithm properly, but the misconceptions return shortly thereafter (Hestenes, 1992). This has been demonstrated often in recent research especially by Hestenes, et. al., and their use of the Force Concept Inventory (Hestenes, 1992). This assessment has been given to students of many different ability levels who have successfully completed introductory physics courses in high school or college that have been taught in a variety of ways (Hestenes, 1987). The results have indicated that traditional lecture instruction is less successful at getting students through their misconceptions than approaches that help students to confront the misconception (Hestenes, 1992, Hake, 1998).

One rationale that has been offered for this discrepancy between success in a physics class and understanding of physics has been offered in the form of cognitive mapping

(Jackendoff, 1999). Those students who primarily use algebraic methods to solve problems do so by first assigning a symbol to each concept. This assignment of a variable actually limits the ability of most students to think about the concept as it changes in time and space. Unfortunately, the concepts emphasized in kinematics are regularly changing in time and space. Students who primarily utilize algebraic methods to solve problems have knowledge of the inputs and outputs of their equations, but they tend to lose the spatial and temporal information that occurred between these two points (Megowan, 2007). As this information is lost, much of the meaning is lost as well. Therefore, the choice of an algebraic method first for solving physics problems can limit the ability of many students to develop a full conceptual understanding of these ideas (Megowan, 2007).

There is nothing inherent in the physics concepts themselves that requires an algebraic method as the primary one for exploration and solution. It is clear that graphs are able to represent a great number of physics concepts, and therefore, the ability to interpret graphs and draw graphs is critical to success in understanding physics (McDermott, 1986). Some go so far as to say that the ability to utilize scientific representations, such as graphing, is the primary difference between good and poor problem solvers (Beichner, 1994). It is also clear that giving students many opportunities over time to work with and interpret graphs improves their skills with graphs (Beichner, 1994, and Mcdermott, 1986). But the great benefit in using graphing to represent physics concepts is that graphing allows one to think multidimensionally and spatially about a problem and understand the nuances of the variables (Jackendoff, 1999).

If students first explore ideas with graphs and diagrams, they should be able to take advantage of their abilities to think spatially (Jackendoff, 1999). Students who utilize this progression have been found to have the ability to think about physical concepts and "to move back and forth in space and time, examine instants or intervals, and see points at which things happened or changed" (Megowan, 2007). Later, the terminology and algebraic representations can be applied to these ideas and students should be able to move between the spatial and algebraic representations without as many difficulties as students who have not gone through the spatial exploration of the concepts (Megowan, 2007).

After establishing that students' misconceptions often get in the way of their ability to retain physics concepts and that traditional teaching methods are not particularly successful in helping students, we have adopted a modeling approach that utilizes multiple representations of physics concepts to help students confront the misconceptions. Graphing has been established as one of the most critical tools to understand physics concepts, and when combined with algebraic representations, results in better student understanding. We will take this progression a step further by first emphasizing the utilization of graphical methods exclusively for problem solving in our classrooms. We expect students will develop the ability to think about kinematics concepts spatially and temporally which will result in a better conceptual understanding of position, displacement, velocity and acceleration. When we follow this with algebraic models of the concepts, students should become adept at translating between these multiple representations of the problem space to solve kinematics problems algebraically with the same or improved capability while simultaneously holding a more accurate conceptual understanding of kinematics.

Method

Subjects:

Investigator 1: Investigator 1 teaches at Highland High School, which is located in Gilbert, Arizona. Gilbert is a suburb of metropolitan Phoenix, and the high school has an enrollment of approximately 3000 students in grades 9-12. The population of the school is approximately 84% Caucasian, 9% Hispanic, 4% Asian, 2% Black, and less than 1% American Indian. Six percent of the students receive free or reduced lunch. Investigator 1 worked with approximately 85 General Physics students. General Physics is populated by about 65 percent seniors and 35 percent juniors. General physics is a college preparatory class.

Investigator 2: Investigator 2 teaches at Cactus Shadows High School in Cave Creek, Arizona, north of Scottsdale. Cave Creek is a suburb of metropolitan Phoenix, and the high school has an enrollment of approximately 1500 students in grades 9-12. The population of the school is approximately 90% Caucasian, 7% Hispanic, 2% Asian, and 1% Black and American Indian. Six percent of the students receive free or reduced lunch. Investigator 2 worked with approximately 85 general physics and honors students. The math prerequisite for general physics is Algebra 2.

Investigator 3: Investigator 3 teaches at Spring Lake Park High School, which is located in a suburb of Minneapolis, Minnesota. Spring Lake Park is one of the smallest suburbs and the high school has an enrollment of approximately 1300 students in grades 9-12. The population of the school is approximately 80% Caucasian, 7% Asian, 6% Black, 5% Hispanic and 2% American Indian. Twenty-six percent of the students receive free or reduced lunch. Investigator 3 worked with approximately 35 Honors Physics students and 70 General Physics students. Honors Physics is primarily populated by juniors and is taught at the level of college freshman level algebra based physics. General Physics is primarily populated by seniors and is a college preparatory class.

Investigator 4: Investigator 4 teaches at Apollo High School, which is located in Glendale, Arizona. Glendale is a suburb of metropolitan Phoenix, and the high school has an enrollment of approximately 1800 students in grades 9-12. The population of the school is approximately 40% Caucasian, 44% Hispanic, 4% Asian, 9% Black, and less than 3% American Indian. 39 percent of the students receive free or reduced lunch. Investigator 4 worked with approximately 10 General Physics students and 10 Honors Physics students. General Physics is populated by about 30 percent seniors and 70 percent juniors. AP Physics is populated by about 90 percent seniors and 10 percent juniors. Both physics classes are college preparatory classes.

Control Group 1: Control group 1 consists of 2 classes of General Physics students taught at Desert Ridge High School in Mesa, Arizona. Mesa is a suburb of metropolitan Phoenix, and the high school has an enrollment of approximately 2400 students in grades 9-12. The population of the school is approximately 71% Caucasian, 17% Hispanic, 3% Asian, 7% Black, and less than 1% American Indian. About 16 percent of the students receive free or reduced lunch. The control group consisted of about 40 General Physics students. General Physics is populated by about 65 percent seniors and 35 percent juniors. General physics is a college preparatory class.

These classes are taught using the modeling curriculum designed, in large part, at Arizona State University. This curriculum allows students to develop models and construct understanding as guided by the instructor.

Control Group 2: Control group 2 consists of one class of General Physics students taught at Spring Lake Park High School, which is located in a suburb of Minneapolis, Minnesota. This is the same school in which Investigator 3 teaches, but this group is not taught by Investigator 3. Spring Lake Park is one of the smallest suburbs and the high school has an enrollment of approximately 1300 students in grades 9-12. The population of the school is approximately 80% Caucasian, 7% Asian, 6% Black, 5% Hispanic and 2% American Indian. Twenty-six percent of the students receive free or reduced lunch. This group consists of about 30 General Physics students. General Physics is primarily populated by seniors and is a college preparatory class. These classes are taught using the modeling curriculum designed, in large part, at Arizona State University. This curriculum allows students to develop models and construct understanding as guided by the instructor.

Procedure for Treatment:

1. Pre-assessment of student abilities.

During the first week of school, we determined the baseline ability level of our students in three separate areas. First, we gave the Purdue Spatial Visualization Test. This helped us determine the relative spatial thinking abilities of our students prior to our treatment. Secondly, we gave the TUG-K test to assess our student's abilities to interpret graphs, move between graphs, and do calculations from graphs. Finally, we utilized the FCI to assess all of the students' understanding of the fundamental ideas of forces.

2. Permission.

No student who is 18 years of age or older was included in the study without their authorization provided by their signature. No student who is 18 years of age or older was photographed or videotaped without their authorization provided by their signature. If students were not 18 years old, then permission to participate in the study, to be photographed, or to be videotaped was gathered from the parent or legal guardian of the student. If permission was not granted to include a student in the study, the results of any assessments were not included in the study. If permission was not granted for photography or videotaping, then those students were not photographed or videotaped.

3. Unit 1

Each investigator completed their own first unit of study to prepare their students for studying physics. This was necessary because of the different backgrounds of the students and their readiness to study physics.

4. Treatment.

This consisted primarily of a change in emphasis from the traditional modeling approach. The modeling curriculum has been designed, in large part, at Arizona State University. This curriculum allows students to develop models and construct understanding as guided by the instructor. Our emphasis was on requiring students to solve problems graphically ***prior*** to developing their ability to solve problems algebraically. The modeling curriculum in physics is commonly used within each of the investigators' districts. The modifications of the curriculum for the treatment were reasonable modifications and within the investigators' discretion as teachers. The treatment was incorporated into Units 2 – 6. These units are constant velocity, constant acceleration, equilibrium, forces and projectile motion respectively. Each investigator was allowed to complete the units in a different sequence so long as they held to the philosophy that graphing methods would be utilized prior to algebraic methods when each new concept was introduced. Two new kinematics worksheets that required students to solve problems graphically were developed for units 2 and 3. Investigators were also able to supplement the modeling materials as long as they held to the philosophy that graphing methods were utilized prior to algebraic methods when each new concept is introduced. The general method for working through each unit was as follows:

a. We introduced each major concept with the paradigm lab for the unit as usual within the modeling curriculum.

b. For a few to several days, depending on the unit, the students solely approached problems from a graphical standpoint. Students then were introduced to algebraic methods to solve the same types of problems. Students finished the unit using graphical and algebraic methods to solve problems.

c. Review for each unit included a worksheet that allowed students to work out each problem with whichever method that they chose, but during the whiteboarding process, both methods were discussed.

d. Each unit ended with a summative assessment of the unit of study.

5. Assessment.

The students were assessed in six ways. First, the students took the TUG-K at the end of unit 3, constant acceleration. Second, the final free-response problems from the unit 2 test and the unit 3 test were kept for investigators' review. Third, each investigator interviewed several students after unit 6 with a challenging problem that can be solved graphically or algebraically utilizing a think-aloud protocol. The questions that were asked of the students in the interview were: 1) What were you thinking when you wrote that?; Can you explain how you solved that?; Why did you use that method to solve your problem?; What approach to the problem was more useful to you?; What does that mean?; What would the graph for this situation look like?; and Can you solve that problem another way?. Fourth, student whiteboards were photographed and some student worksheets and tests were kept from all groups so that the progress of students could be monitored. Each student took the FCI after completing circular motion, the unit following our treatment. Each of the assessments given was a common assessment in physics that is used in the investigators' districts and other districts around the country. Each investigator kept field notes on details of our study to supplement our qualitative data.

Modifications from the Proposal.

We had originally intended to utilize a free-response question from the end of the forces unit (unit 6); however, the question we used was too complex to analyze qualitatively because it had too many parts. We therefore excluded that question from our analysis.

Only one of the investigators gave a third TUG-K test to their students because of the time pressures of the school year. For the same reason, only one of the investigators gave the FCI for a third time.

One of the investigators gave the Purdue Spatial Inventory for a second time.

Data Analysis

In this study, we collected both quantitative and qualitative data. The quantitative data came from the Purdue Spatial Visualization Test, the TUG-K and the FCI pretests, and the TUG-K and FCI posttests. The qualitative data came from examining student whiteboards discussions during class (specifically during the review worksheets), collected worksheets, test questions as well as student interviews after the completion of the treatment and the investigator's field notes. The data analysis consisted of statistical analysis of scores and an analysis of the qualitative artifacts gathered to look for patterns and correlations with whiteboard representations and interview remarks.

We examined our quantitative data in the following ways:

1. We compared the results of the students who received the treatment on the TUG-K to the results of the students in the control group on the TUG-K. This allowed us to determine if our treatment helped the students understand kinematics conceptually.

2. We compared the FCI scores on questions 8, 10, 12, 14, 17, 19, 20, 21, 22, 23, 24 (hereinafter "selected FCI questions") of the investigators' students to the FCI scores of the of the control group on the same questions to determine if our treatment improved students' ability to analyze physics problems that were particularly spatial in nature.

3. We analyzed the data from the control group to the group that received our treatment of the last free-response question in units 2 and 3 to determine the number of students in each group who accurately solved the problem, and to determine the proportion of students in each group who utilized graphing in formulating their solution. We also analyzed this data qualitatively as described below.

4. We determined what, if any, correlation there was between students who scored well on the Purdue Spatial Inventory and students who responded to our treatment.

We examined our qualitative data in the following ways:

1. We compared the data from the control group to the group that received our treatment of the last free-response question in units 2 and 3 to determine if there was a qualitative difference in the overall level of problem solving ability between the two groups and to determine what effect the utilization of a graph helped the students to solve the problems accurately.

2. We interviewed a number of students who received our treatment utilizing a think-aloud protocol to investigate if they chose graphing first to solve problems and to determine how the treatment affected the student's abilities to think conceptually about kinematics concepts.

Timeline

1. Treatment

The treatment was performed from units 2 through unit 6 which took place from about the third week of the school year until near the end of the first semester. The primary emphasis was on utilizing a graphing first approach to solving problems. Also, students were asked to supply graphical reasoning throughout the equilibrium, forces and projectile motion units where they have not otherwise been asked to do so before. This was done by incorporating additional worksheets, additional graphs on worksheets, and adding graphs to whiteboard explanations to supplement the typical modeling curriculum.

2. Assessment

As described above, the Purdue Spatial Visualization Test, the TUG-K and the FCI were given to our students during the first week of the school year as pre-assessments. The TUG-K was given to the students by each investigator at the end of unit 3 on constant acceleration. The FCI was given again shortly after the completion of circular motion (unit 7) which did not receive our treatment. Student interviews were conducted after the completion of the last unit of the treatment. As noted in the procedure, the last unit may not have been unit 6 depending on the order the investigator chose to present the units. In addition, the last free-response question on the unit 2 and 3 tests were saved from each student receiving the treatment as well as each member of the control group.

During the student interviews, each student was individually presented with a problem that they were asked to solve while describing their thinking process verbally. All questions asked by the investigators: 1) were focused on the physics problem at hand, 2) helped elicit the student's thinking process, 3) determined what additional information a student might need to complete the problem.

Results

The statistical tests that we performed prior to reporting our results clearly demonstrated that the students who self-selected themselves, primarily by their own evaluation of their readiness for physics and willingness to work hard, into the honors classes responded much better overall to our treatment than the students in general physics who had the same teachers responded to our treatment. We therefore determined that it would be more appropriate to separate the honors physics students from the general physics students when comparing the "treatment" group to the control group. Those investigators who taught honors physics analyzed that portion of their data separately and reported in their own sections of this paper; and the details are included therein. Our decision was based primarily on our interest in not misevaluating our treatment because of skewed data from honors students.

TUG- K

We were interested in the effect of the treatment on students' kinematics understanding as measured by the TUG-K. Based on a non-directional paired samples t-test at alpha = 0.05, we reject the null hypothesis that population mean of the pretest takers is equal to the population mean of the posttest takers, $t(168) = 23.6$, $p < 0.001$. We conclude that there was a significant difference in kinematics understanding as measured on the TUG-K for individuals before the treatment (Mean = 4.25, SD = 2.425) and after the treatment (Mean = 11.36, SD = 4.201). We are 95 % confident that the interval 7.702 and 6.511 contains the true population mean difference. The correlation was 0.401. (1 in Appendix A)

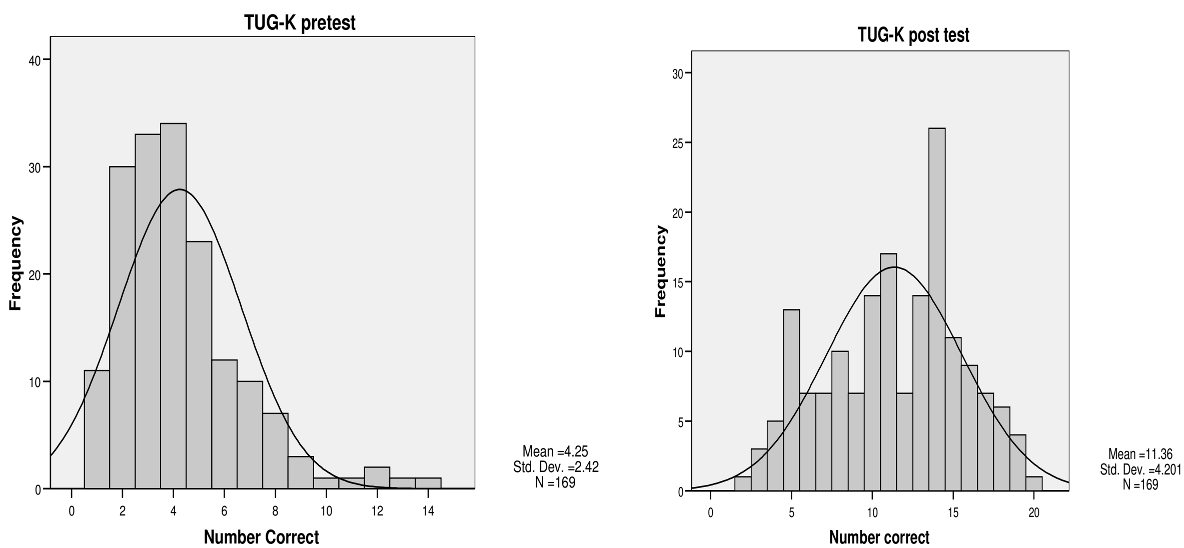


Figure 1. Comparison of histograms for treatment TUG-K pretests and posttests.

The t-test confirms what appears to be evident by looking at the histograms. Initially, the scores are skewed towards the low performance side of the graph. No student in the treatment group scored above a 14 on the posttest. The posttest histogram is nearly Gaussian and very slightly skewed towards the high performance side of the graph. Ten students scored 18 or above on the pretest. These two groups of tests appear to have come from different populations of students, and the statistical test supports that. We conclude that the students who received our treatment showed significant gains in their understanding of kinematics concepts.

The intention of the investigators was to compare the gains of the treatment group to a control group. To further this goal, we were interested in determining whether the treatment samples and the control samples from the TUG-K pretest were from the same population. Based on a non-directional independent samples t-test at alpha = 0.05, we reject the null hypothesis that population mean of the treatment group is equal to the population mean of the control group, $t(43.612) = 6.47$, $p < 0.001$. We conclude that there was a significant difference between the treatment sample (Mean = 4.25, SD = 2.420) and the control sample (Mean = 9.40, SD = 4.893) on the TUG-K pretest. We are 95 % confident that the interval 3.547 to 6.755 contains the true population mean difference. (2 in Appendix A)

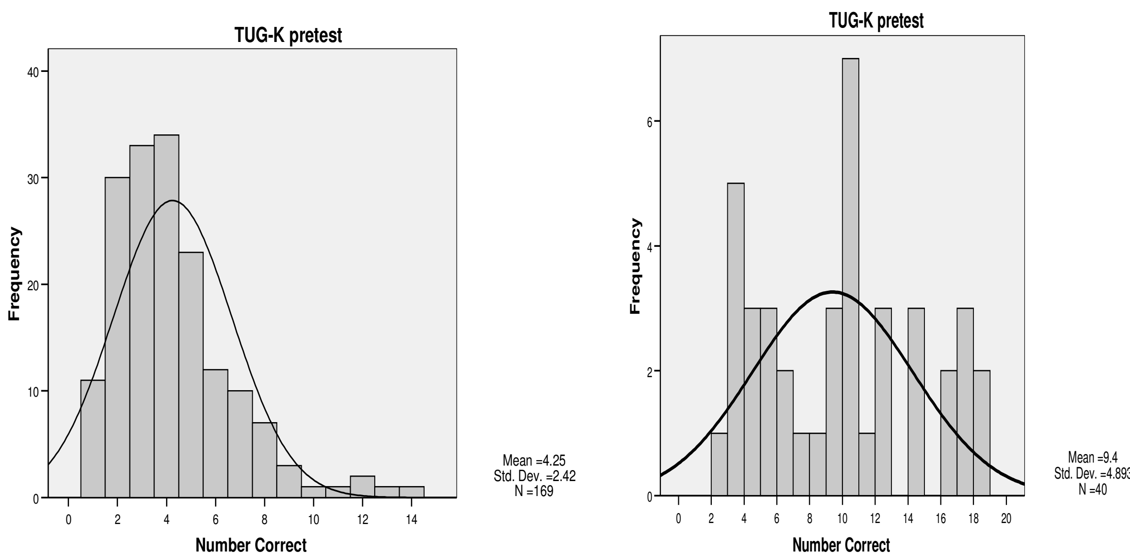


Figure 2. Treatment TUG-K pretest histogram on the left. Control TUG-K pretest histogram on the right.

Inspection of these two histograms shows that the first is skewed towards low performance while the second is fairly evenly distributed. The means of the two populations are separated by about 6 on a test with only 20 questions. It is apparent that these two groups of test scores come from different populations of students. The inspection of the graphs confirms the t-test that these two groups of students are not from the same population.

It is worth noting that this result was surprising to the investigators because the control group came from a school that is very similar in many aspects from the group of investigator 1. Investigator 1 had the most students represented in the treatment group. The schools are in the same district, only a few miles apart, and have a similar demographic and socioeconomic make-up. We spoke with the control group teacher at length after finding these results and have determined a possible reason for the difference between the samples.

The school for the control group has 8 physics classes from 2400 overall students in grades 9 – 12, and 5 of the 8 classes are general physics. Only the very confident and well-accomplished students tend to take physics at all, and often they choose to take regular physics even if they are capable of being in an honors class. Investigator 1's school has 14 physics classes for about 3000 overall students in grades 9-12, and 8 of the 14 classes. This school has a science registration day in the spring that is designed to convince the students to take the highest level of science class that they are prepared to take. That results in most of the higher level students going to honors physics and AP physics with mostly the lower level, but still capable, students to populate the general physics classes. We believe that this difference in school culture explains why the treatment and control groups were found to come from different populations on the TUG-K pretest.

Because our treatment and control groups had not come from the same population, we looked for methods other than t-tests that we could use to compare these groups. We learned of the Hake test for comparing the gains made between differing samples.

The Hake test was developed by Richard R. Hake from the University of Indiana, and he used it in at least one case to compare course effectiveness in promoting conceptual understanding of mechanics and forces as measured by pretest and posttest data from the Mechanics Baseline Test and the FCI (Hake, 1998).

The Hake test allows for a comparison of gains between samples that have taken pretests and posttests on the same instrument. The formula for computing the Hake gain for an individual student is:

$$\frac{\text{posttest score} - \text{pretest score}}{\text{total possible} - \text{pretest score}}$$

Essentially, this is a ratio of an individual student's gain from pretest to posttest to their total possible gain. When evaluating the results, a Hake score of 1 indicates a maximum student gain, and a Hake score of 0 indicates no gain for the student. Negative Hake scores are possible if a student receives a lower score on the posttest than on the pretest. The Hake scores for all of the individuals in a sample can be averaged for the entire sample to determine an average sample Hake score, and this is the test that we performed on our samples. A score of $\langle g \rangle_{NP} = M \pm SD$ describes "an average Hake gain for a sample size of N, an identifier of P (i.e. T for treatment or C for control), a mean of M and a standard deviation of SD." "High" gains are defined as (<g>) $\geq$ 0.7; "medium" gains are defined as 0.7 > (<g>) $\geq$ 0.3; "low" gain scores are defined as (<g>) < 0.3 (Hake, 1998).

We were interested in comparing the Hake gains of the treatment group to the Hake gains of the control group on the TUG-K. Based on an average Hake gains test, we determined that there was a substantially larger gain for the treatment group $\langle g \rangle_{169T} = 0.46 \pm 0.25$, than for the control group $\langle g \rangle_{40C} = 0.24 \pm 0.51$. (3 in Appendix A)

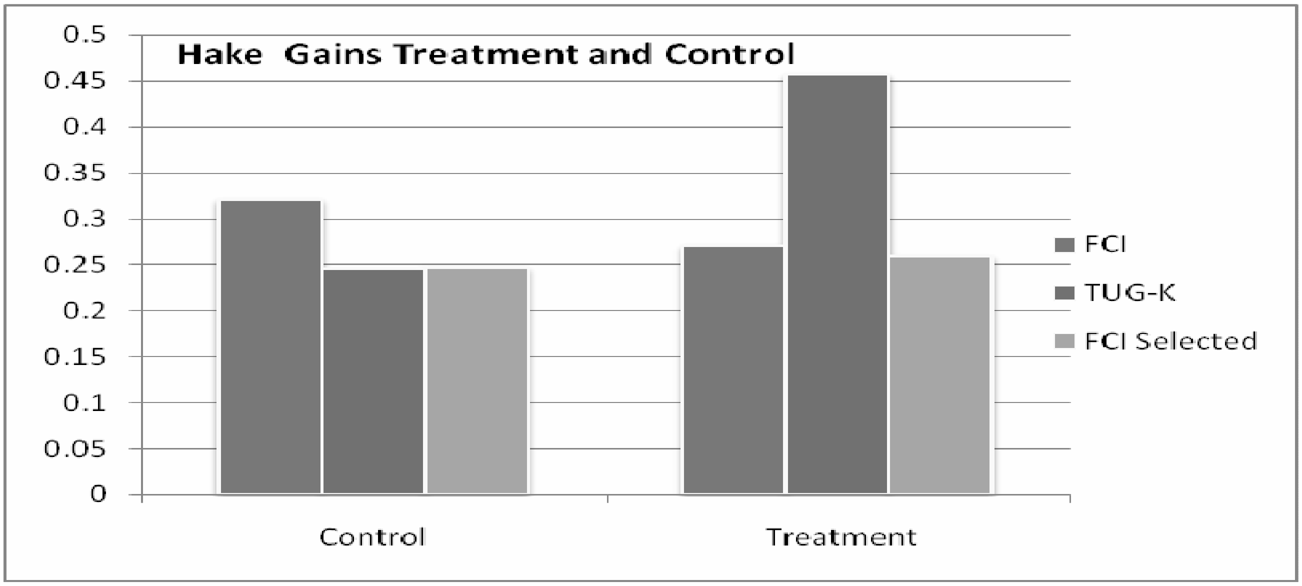


Figure 3. Histogram showing the average Hake gains, <g>, for the treatment and control groups on the FCI, TUG-K and FCI Selected questions.

One method of determining the significance of an average Hake gain is to look at the number of standard deviations between average Hake gain of the treatment group and the average Hake gain of the control group. This is a calculation that has been done by Hake himself (Hake, 1998). This calculation can be performed as follows:

$$\frac{\text{Treatment } \langle g \rangle - \text{Control } \langle g \rangle}{\text{Standard Deviation of Treatment } \langle g \rangle}$$

Essentially, this calculation tells you the number of treatment standard deviations that the treatment group scored above the control group.

The average Hake gain for the treatment group was 0.88 standard deviations above the average Hake gain of the control group. In addition, our Hake gain scores were in the medium gain range while the control group Hake gain fell in the low range. Therefore, we believe that our treatment resulted in greater gains in our students' kinematics understanding than the control group. A determination of this result was the primary goal and hope of the research team.

There were two difficulties at this point regarding the analysis of the Hake gain data. First, we could not find any definitive reference regarding the number of standard deviations for the treatment group scores above the control group scores that indicates significance. Secondly, we realized that the standard deviation of the control group was not a part of the calculation above, and we were concerned about the size of the standard deviation of the control group's score (0.51). This appeared to be very large compared to the range of probable Hake gain scores of 0 to 1. So, we conducted a t-test to determine whether or not the average Hake gains could have come from the same population.

Based on a non-directional independent samples t-test at alpha = 0.05, we reject the null hypothesis that population mean of the treatment group Hake gain is equal to the population mean of the control group Hake gain, $t(43.42) = -2.524$, $p = 0.015$. We conclude that there was a significant difference between the treatment sample average Hake gain (Mean =0 .458, SD = 0.250) and the control sample average Hake gain (Mean = 0.24468, SD = 0.517) on the TUG-K pretest. We are 95 % confident that the interval -0.3815 to -0.0426 contains the true population mean difference. (4 in Appendix A)

This t-test tells us that the average Hake gains do not come from the same population. Each of our three indicators (the treatment gain being 0.88 standard deviations above the control gain, the treatment "high" gain compared to the control "low" gain, and the t-test showing the scores are from different populations) support our interpretation that the difference in Hake gain scores between the treatment and control group is significant.

Qualitative Data

We decided to look at some of the qualitative data to see if it also supported the conclusion that our treatment improved students' understanding of kinematics more than just modeling alone. Our qualitative data primarily consisted of two strip questions (so named because they were given on vertical half-sheets of paper). One question was from the end of unit 2 on constant velocity. (5 in Appendix A) This question asked students to find the total displacement of an object that traveled at 3 different constant velocities for 3 different time periods. The investigators consider this problem to be a straightforward and reasonable kinematics problem for high school students to solve. The second strip question was given at the end of unit 3, and it asked students to find the total displacement of an object that traveled at a constant velocity for a short period then accelerated for a short time until it stopped. A second portion of the question asked the students to determine the acceleration of the object only for the time that it was changing velocity. (6 in Appendix A) The investigators consider this problem to be a fairly challenging kinematics problem for a general physics student in high school. Both questions were given to the treatment group and the control group. Both of these questions were written so that either a graphical method or an algebraic method could be used to solve them.

Each of the strip questions was categorized based on the method that the students used to solve the problem (solely graphical, solely algebraic, or a combination of graphical and algebraic method). We had originally intended to further categorize each question as being correct or incorrect. However, the students had a great deal of difficulty interpreting the direction of travel for one part of the trip of the first strip question, and each investigator treated the situation with their students differently. Therefore, only the second strip question was evaluated as being "essentially correct" (entirely correct or with one minor error) or "incorrect."

The graphs below show the differing methods of solutions chosen for the treatment and control groups on the first strip question. We had originally expected that the treatment group would choose graphing at a much higher rate than the control group. That expectation was not met for the first strip question.

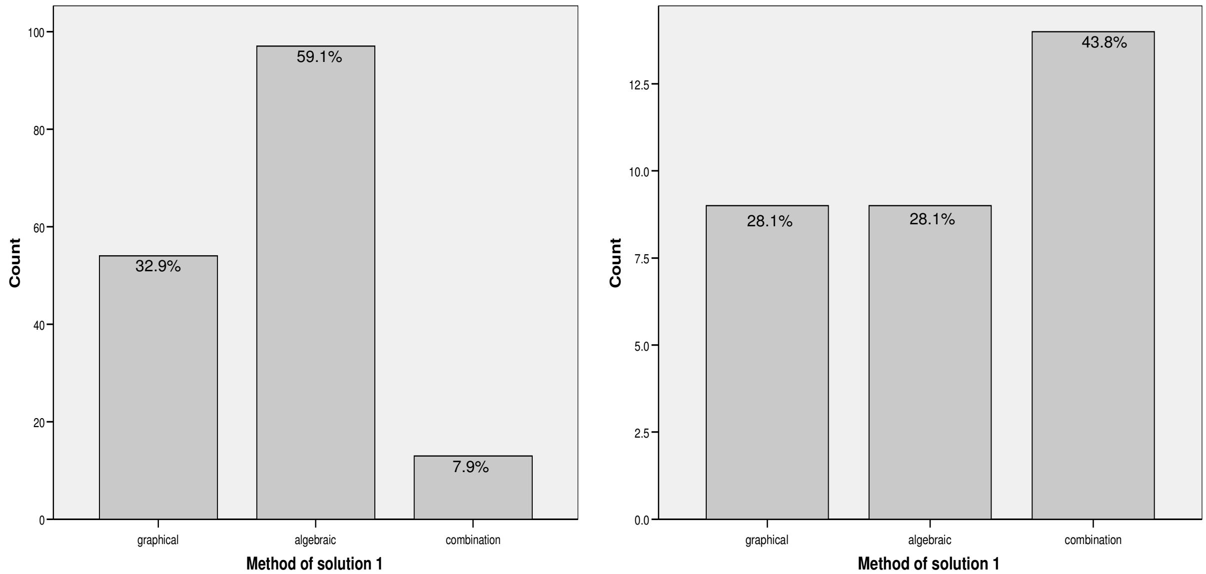
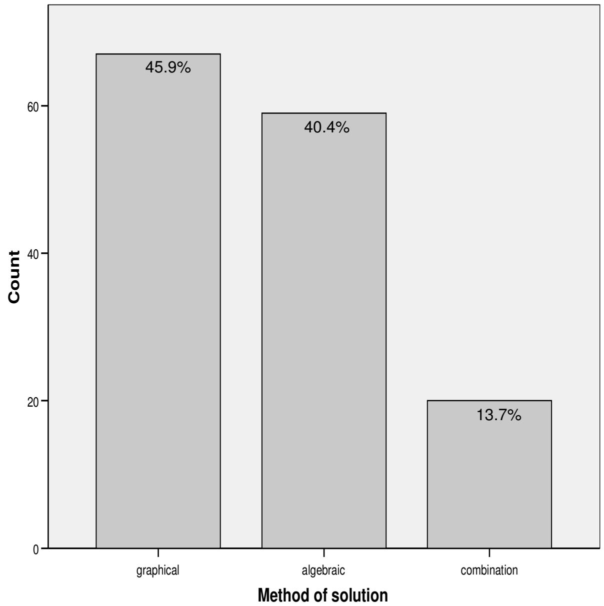


Figure 4. Histograms of the method of solution for strip question 1. Treatment questions are on the left, and control questions are on the right.

In the first strip question, 32.9 percent of the students in the treatment group solved the problem solely by graphing, and 28.1 percent of the control group did the same. However, when the students who solved the problem with a combination of methods are included, 41.8 percent of the treatment group and 71.9 percent of the control group used graphing for at least a component of their solution.

We then analyzed the strip questions from unit 3 in the same manner, and we were struck by the distinct differences. The graphs below show the differing methods of solutions chosen for the treatment and control groups on the second strip question. It is easy to see from the graphs that for the second question, the students in the treatment group chose graphing at a much higher rate to solve this problem than the control group. This does in fact match with what the investigators had predicted for a result. We find that 45.9 percent of the students in the treatment group solved the problem solely by graphing, while only 9.8 percent of the control group did the same. Furthermore, when the students who solved the problem with a combination of methods are included, 59 .6 percent of the treatment group and only 9.8 percent of the control group used graphing for at least a component of their solution. None of the control group students used a combination of methods to solve this problem.



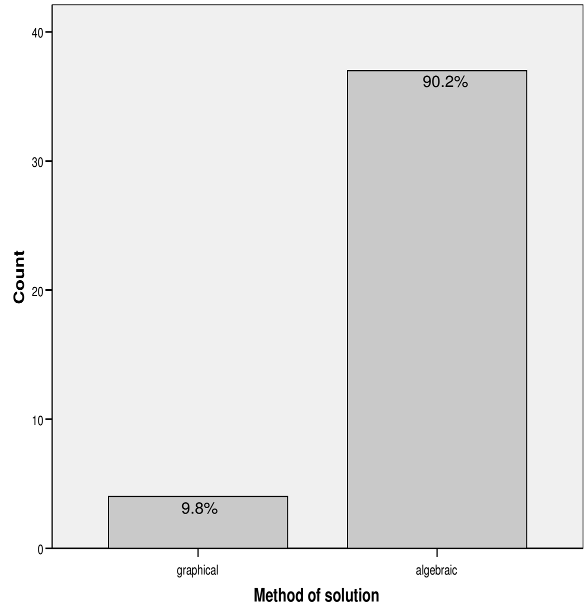


Figure 5. Histograms of the method of solution for strip question 2. Treatment questions are on the left, and control questions are on the right. There were no combination solutions for the control group

Because the second strip question was given to the students at the end of the kinematics unit, we believe that it is a better indicator of the results of the treatment than the first strip question that was given an entire unit earlier. We conclude that by the end of the treatment the students who received the treatment chose graphing more frequently to solve problems than students who did not receive the treatment.

The investigators have concluded, however, that the students' use of graphing to help solve the problem is not meaningful if they ultimately do not solve the problem accurately. In other words, we wanted the students to get the problems right regardless of how they did it. The graphs below show the proportion of correct and incorrect solutions relative to the method of solution chosen by each student.

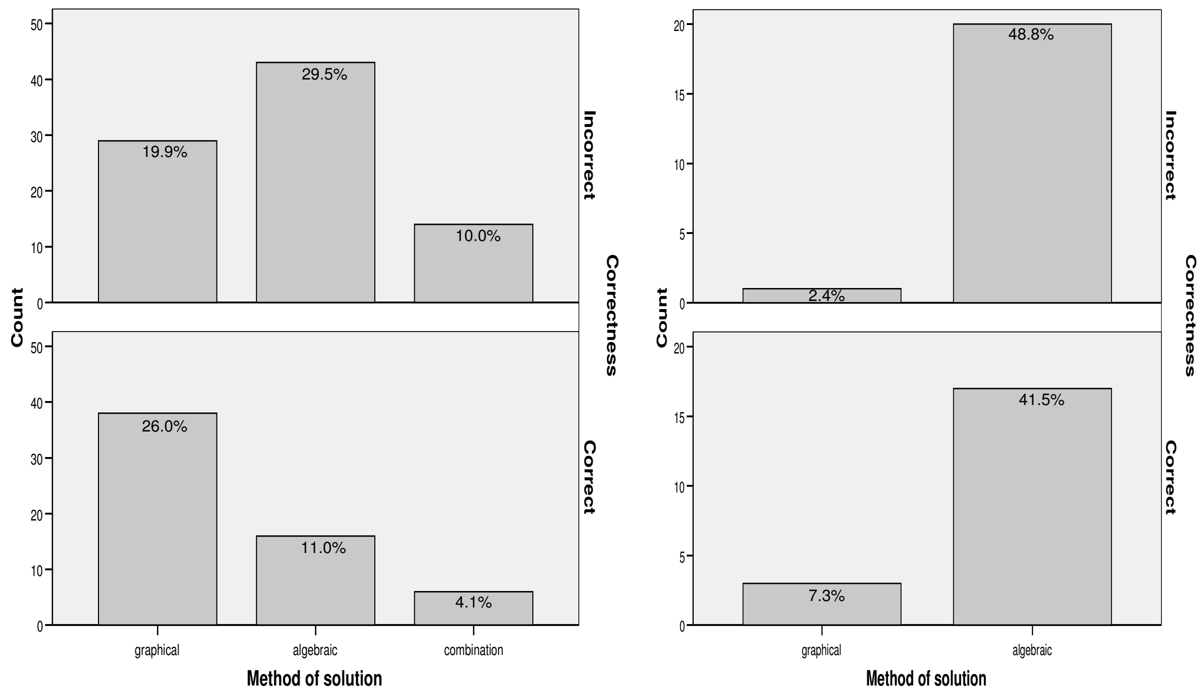


Figure 6. Graphs of accuracy of solution for strip question two broken out by method of solution. Results of the treatment group are on the left. Results of the control group are on the right.

Inspection of these graphs shows that the students who received the treatment and graphed at least a portion of their solution were accurate 50.6 percent of the time on the second strip question. Those who solved algebraically were accurate only 27.1 percent of the time. The students who solved the problems algebraically in the control group were accurate 45.9 percent of the time. Those who solved graphically in the control group were accurate 75 percent of the time, but there were only four of these students overall, so we discount these results.

We conclude from the qualitative data that we were successful in getting our students to utilize graphing to solve a complicated kinematics problem, and this did not decrease their overall level of kinematics problem solving. In fact, those of our students who chose to graph were more successful in solving the problem than those who only solved algebraically.

FCI

We were interested in the effect of the treatment on students' understanding of force concepts as measured by the FCI. Based on a non-directional paired samples t-test at alpha = .05, we reject the null hypothesis that population mean of the pretest takers is equal to the population mean of the posttest takers, $t(153) = -17.119$, $p < 0.001$. We conclude that there was a significant difference in students' understanding of force concepts as measured on the FCI for individual students before the treatment (Mean = 6.47, SD = 2.542) and after the treatment (Mean = 12.82, SD = 4.915). We are 95 % confident that the interval -7.086 and -5.620 contains the true population mean difference. The correlation was 0.382. (7 in Appendix A)

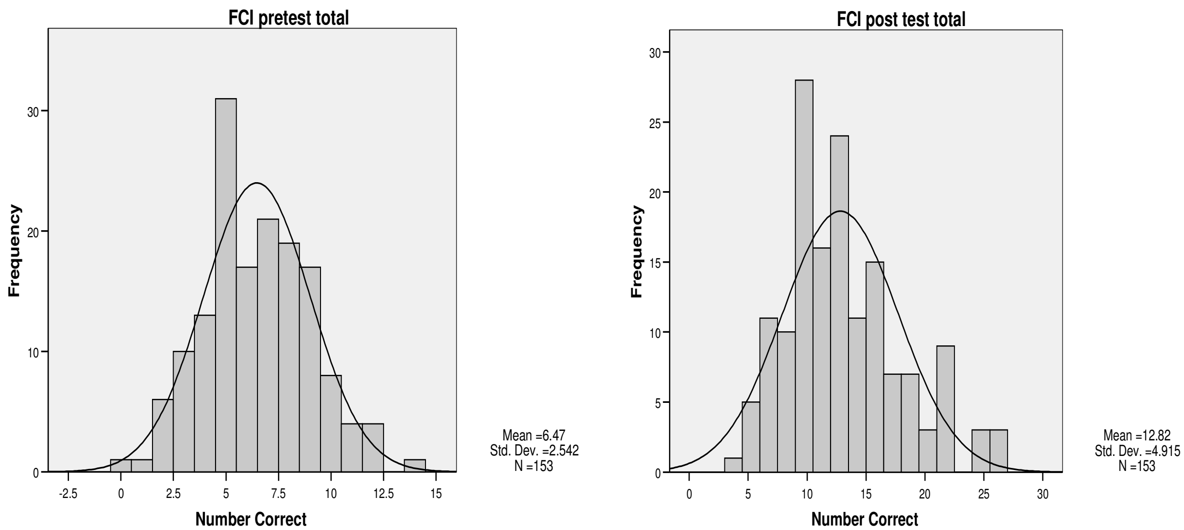


Figure 7. Comparison of histograms for treatment FCI pretests and posttests.

From these results, we conclude that the students who received our treatment showed significant gains in their understanding of force concepts.

Our determination that the treatment and control groups came from a different population on the TUG-K required us to look at the same possibility for the FCI. We were interested in determining whether the treatment samples and the control samples from the FCI pretest were from the same population. Based on a non-directional independent samples t-test at alpha = 0.05, we reject the null hypothesis that population mean of the treatment group is equal to the population mean of the control group, $t(44.051) = 2.785$, $p = 0.008$. We conclude that there was a significant difference between the treatment sample (Mean = 6.47, SD = 2.542) and the control sample (Mean = 8.19, SD = 3.504) on the FCI pretest. We are 95 % confident that the interval 0.476 to 2.971 contains the true population mean difference. (8 in Appendix A)

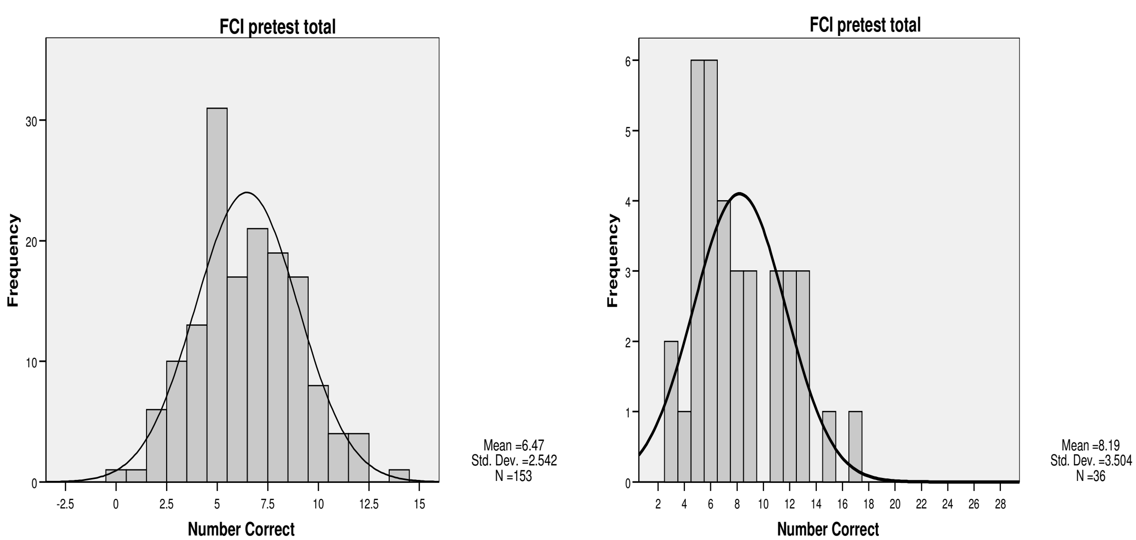


Figure 8. Treatment FCI pretest histogram on the left. Control FCI pretest histogram on the right.

As with the TUG-K results, we conclude that the treatment group and control group who took the FCI pretest did not come from the same population, and a traditional statistical comparison between these samples' scores on the FCI using further t-tests is inappropriate.

We also examined the two groups' performance on the FCI using the Hake test. We were interested in comparing the Hake gains of the treatment group to the Hake gains of the control group on the FCI to determine if the treatment produced statistically larger gains. Based on an average Hake gains test, we determined that there was not a substantially larger gain for the treatment group $\langle g \rangle_{153T} = 0.27 \pm 0.20$, than for the control group $\langle g \rangle_{36C} = 0.32 \pm 0.18$. (9 in Appendix A) Because the treatment group's average Hake gain was 0.28 standard deviations below the Hake gain of the control group on the FCI, we conclude that our treatment did not improve students' understanding of force concepts more than a standard modeling approach.

Selected FCI questions

In addition to examining the overall FCI scores, we selected eleven questions from the FCI that we believed a change in spatial reasoning ability would have the most effect upon. We were interested in the effect of the treatment on students' understanding of particularly spatial force concepts as measured by the Selected FCI questions. Based on a non-directional paired samples t-test at alpha = 0.05, we reject the null hypothesis that population mean of the pretest takers is equal to the population mean of the posttest takers, $t(153) = -12.178$, $p < 0.001$. We conclude that there was a significant difference in students' understanding of particularly spatial force concepts as measured on the Selected FCI questions for individual students before the treatment (Mean = 2.68, SD = 1.669) and after the treatment (Mean = 4.92, SD = 2.362). We are 95 % confident that the interval -2.598 and -1.873 contains the true population mean difference. The correlation was 0.407. (10 in Appendix A)

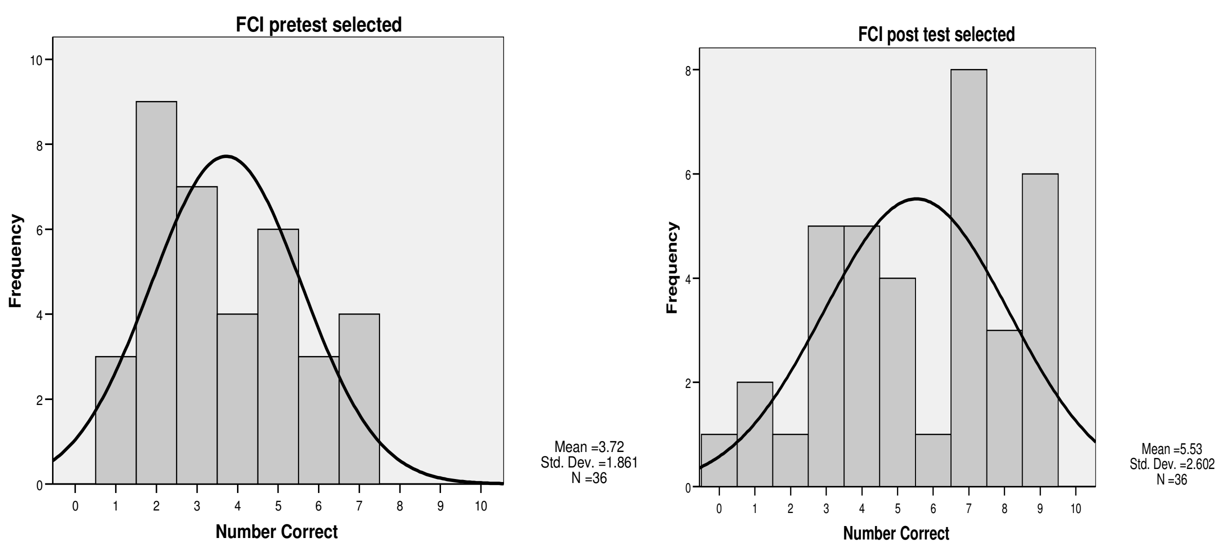


Figure 9. Comparison of histograms for treatment Selected FCI pretests and posttests.

By looking at the graphs, we see that the pretest scores were skewed towards low scores, and the posttest scores were pretty normally distributed. This, in combination with the t-test allow us to conclude that the students who received our treatment showed significant gains in their ability to analyze spatial force concepts.

Next, we examined the treatment and control groups with regards to their performance on the pretest selected FCI questions. We were interested in determining whether the treatment samples and the control samples from the selected FCI pretest were from the same population. Based on a non-directional independent samples t-test at alpha = 0.05, we reject the null hypothesis that population mean of the treatment group is equal to the population mean of the control group for the pretest selected FCI questions, $t(187) = 3.298$, $p = 0.001$. We conclude that there was a significant difference between the treatment sample (Mean = 2.68, SD = 1.669) and the control sample (Mean = 3.72, SD = 1.861) on the pretest selected FCI questions. We are 95 % confident that the interval 0.419 to 1.666 contains the true population mean difference. (11 in Appendix A)

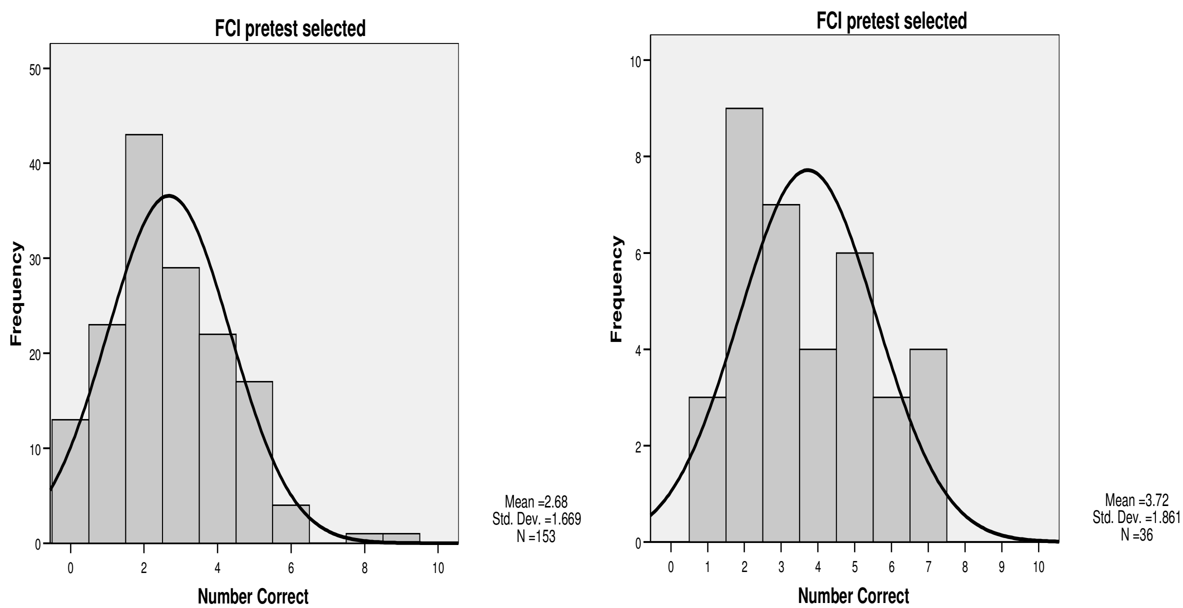


Figure 10. Treatment pretest Selected FCI questions histogram on the left. Control pretest selected FCI histogram on the right.

From these results, we conclude that the treatment and control groups who took the pretest selected FCI questions did not come from the same population, and a traditional statistical comparison between these samples' scores on this test is inappropriate.

Finally, we examined the two groups' performance on the selected FCI questions using the Hake test. We were interested in comparing the Hake gains of the treatment group to the Hake gains of the control group on the selected FCI questions to determine if the treatment produced statistically larger gains. Based on an average Hake gains test, we determined that there was not a substantially larger gain for the treatment group $\langle g \rangle_{153T} = 0.26 \pm 0.31$, than for the control group $\langle g \rangle_{36C} = 0.25 \pm 0.30$. (3 in Appendix A) Because the treatment group's average Hake gain was less than 0.1 standard deviations above the Hake gain of the control group on the selected FCI questions, we conclude that our treatment did not improve students' understanding of spatial force concepts more than a standard modeling approach.

Correlations with the Purdue Spatial Inventory

The treatment utilized in this study was focused on providing a method for students to think spatially about physics problems as a context for kinematics problem solving. As a result, we were concerned that it would only be those students who entered the class with developed spatial thinking skills that would benefit from our treatment. This thought comes with the caveat, however, that some minimum level of spatial thinking ability is probably a prerequisite to success in a physics class regardless of the method of instruction.

With these ideas in mind, the Purdue Spatial inventory was given the first week of the school year as a method of determining the baseline level of spatial ability of our students. (Appendix B) Correlations were then run between the Purdue results and the pretest and posttest results for the TUG-K, the FCI and the Selected FCI questions.

Correlations

		Purdue Spatial pretest	TUG-K pretest
Purdue Spatial pretest	Pearson Correlation	1	.345**
	Sig. (2-tailed)		.000
	N	191	191
TUG-K pretest	Pearson Correlation	.345**	1
	Sig. (2-tailed)	.000	
	N	191	304

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

Correlations

		Purdue Spatial pretest	TUG-K post test
Purdue Spatial pretest	Pearson Correlation	1	.398**
	Sig. (2-tailed)		.000
	N	191	187
TUG-K post test	Pearson Correlation	.398**	1
	Sig. (2-tailed)	.000	
	N	187	302

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

Figure 11. Correlation results for the Purdue spatial inventory compared to the TUG-K pre- and posttests.

The correlation between the Purdue and the TUG-K pretest was low, 0.345. The correlation between the Purdue and the TUG-K posttest was slightly higher, 0.398, but still in the low range for correlations overall.

Our conclusion is that the level of our students' spatial thinking skills as measured by the Purdue Spatial Inventory played very little (if any) role in our students' response to our treatment.

Conclusion

The research in the field of physics education strongly supports the idea that students who are taught physics with an interactive method learn substantially more physics than those who are taught in a traditional lecture format. The research further indicates that when the interactive methods allow students to think spatially prior to assigning algebraic variables to a concept, they understand them better. Our research has been an attempt to push that idea to its limits by requiring the students to utilize graphical methods for solving kinematics problems before even presenting the possibility of an algebraic method of solution.

Our favorable results occurred in the most likely place in our study, the TUG-K. With this instrument, we were able to see how our treatment affected our students' understanding of kinematics concepts compared to the control group that was taught utilizing a modeling approach. Our students' average Hake gains on the TUG-K were significantly better than those of the control group, and this was supported with a t-test and to some extent the strip questions. We are certain that by the end of our treatment many of our students were comfortable using graphing to solve complicated kinematics concepts, and most of them chose to graph first.

A lot of time and hard work went into looking at the FCI and selected FCI questions for some sign that our treatment had an effect; however, in retrospect, it was probably unlikely that our treatment that focused so strongly on kinematics concepts was going to have a statistically measurable effect on student's FCI responses generally.

There are parts of our research that we would change if we were to do the project over again. First of all, we would be careful to select a larger control group that was more likely to come from the same population as the treatment group. We also considered that this project may have been better suited to a two year approach with one year being taught with modeling and the second being taught by our treatment. This would have assured that each teacher would have a control from their typical students at their school with only the annual variations in sample at the same school.

When we look at the population of high school physics students overall, and the general physics students in particular, we see that there are very few who will leave us and pursue a degree or a career in physics or engineering. While we would love for those numbers to increase, the reality is that what most of our students will gain from a year of physics in high school is better problem solving and data analyzing skills. Our students will not face many kinematics problems, but they will likely face many graphs no matter where they go. Our students are more comfortable making and interpreting graphs after our treatment than they would have been without our treatment, and most of them can use these graphs to solve problems. This is good.

In our minds, we have succeeded with this research project because we demonstrated a positive result in our students that will help them succeed in life regardless of whether they continue studying physics.

Investigator 1 Field Trial Report

Introduction

Investigator 1 is Jim Archambault from Mesa, Arizona. This year was my 5th year of teaching physics. I teach at Highland High School in Gilbert, Arizona. This is a middle class school that has consistently had 90 percent of its students pass the math, reading and writing standardized tests given by the State of Arizona each year. I taught 4 sections of general physics this year, and almost every one of those students and their parents agreed to inclusion in the study.

Procedure

There were only a few minor differences between the implementation of the treatment for investigator 1's group compared to the treatment group overall. These differences mostly revolved around the amount of time available to do the treatment in my district. Because my district has a rigid district final exam at the end of the first semester, it was imperative that I work through kinematics and forces at a fairly brisk pace.

Results

Tug-K

Upon reviewing all of the data from the treatment group overall, I decided to look at how my students fared independent of the rest of the treatment group. The most obvious places to look for gains were the places where the treatment group had success, on the TUG-K.

I was interested in looking for the effect of the treatment on my students' kinematics understanding as measured by the TUG-K. Based on a non-directional paired samples t-test at alpha = 0.05, I reject the null hypothesis that population mean of the pretest takers is equal to the population mean of the posttest takers, $t(84) = 26.421$, $p < 0.001$. I conclude that there was a significant difference in kinematics understanding as measured on the TUG-K for individuals before the treatment (Mean = 4.26, SD = 2.606) and after the treatment (Mean = 13.16, SD = 3.124). I am 95 % confident that the interval 9.576 to 8.236 contains the true population mean difference. The correlation was 0.423. (12 in Appendix A)

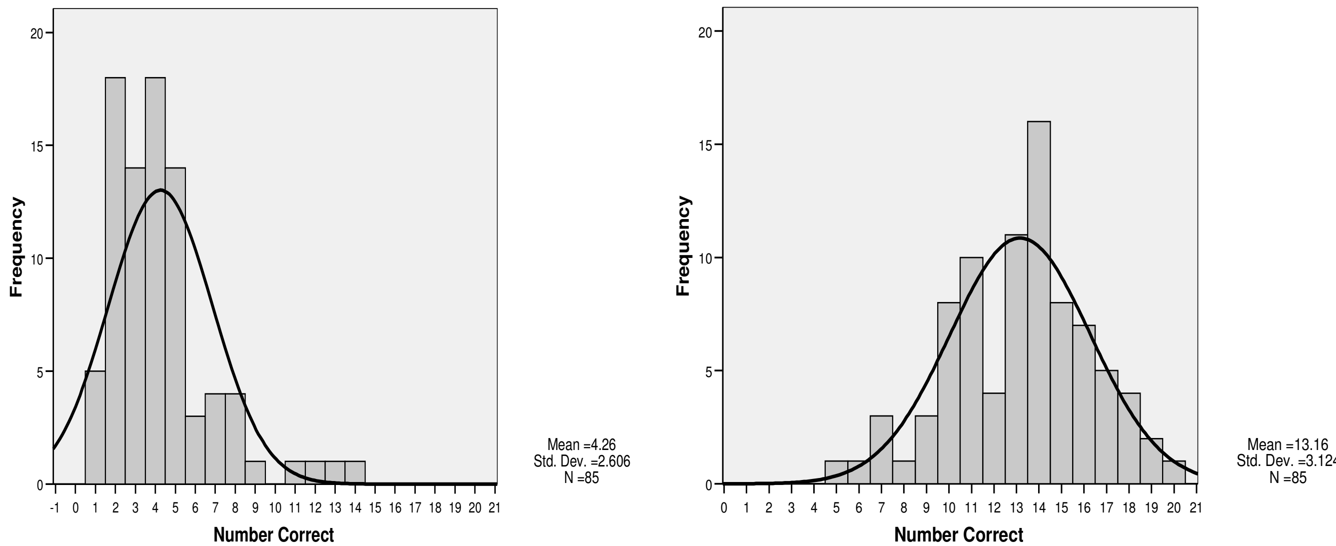


Figure 1. Comparison of histograms for Archambault's TUG-K pretests and posttests.

The histograms above clearly show that the pretest sample was skewed towards low scores, and the posttest sample was skewed towards higher scores, and the mean score increased from 4.26 to 13.16. Because the posttest sample is shown to be from a different population than the pretest sample (in a positive way) I conclude that the students who received the treatment showed significant gains in their understanding of kinematics concepts. Furthermore, I conclude that my sample performed similarly to the treatment group overall.

Despite our previous finding that the treatment group and the control group were not from the same population, I did not know if that was in fact the case for my sample. So I was interested in determining whether my sample and the control sample from the TUG-K pretest were from the same population. Based on a non-directional independent samples t-test at alpha = 0.05, I reject the null hypothesis that population mean of the treatment group is equal to the population mean of the control group, $t(49.694) = 6.242$, $p < 0.001$. I conclude that there was a significant difference between the treatment sample (Mean = 4.26, SD = 2.606) and the control sample (Mean = 9.40, SD = 4.893) on the TUG-K pretest. I am 95 % confident that the interval 3.487 to 6.796 contains the true population mean difference. (13 in Appendix A)

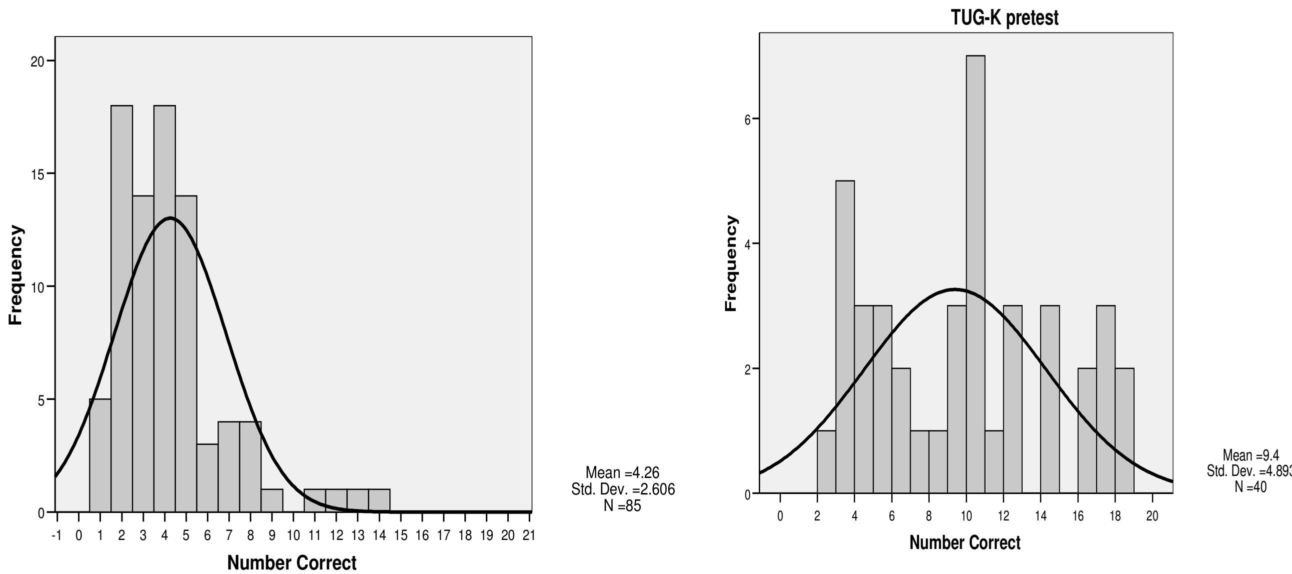


Figure 2. Archambault's TUG-K pretest histogram on the left. Control TUG-K pretest histogram on the right.

Because the t-test reveals that Archambault's group and the control group came from different populations, I conclude that, a traditional statistical comparison between these samples' scores on the TUG-K using further t-tests on this test is inappropriate.

After we had success utilizing the Hake gains for comparing the treatment group to the control group, it seemed appropriate to make the same comparison with the data from Archambault and the control group.

I was interested in comparing the Hake gains of the treatment group to the Hake gains of Archambault's group on the TUG-K. Based on an average Hake gains test, I determined that there was a substantially larger gain for Archambault's group $\langle g \rangle_{85T} = 0.57 \pm 0.18$, than for the control group $\langle g \rangle_{40C} = 0.24 \pm 0.51$. (14 in Appendix A)

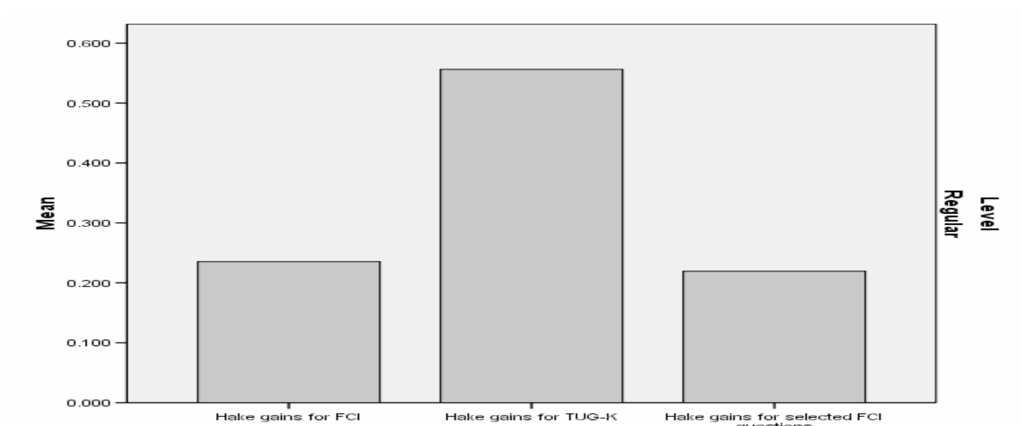


Figure 3. Histogram showing the average Hake gains, <g>, for Archambault's group.

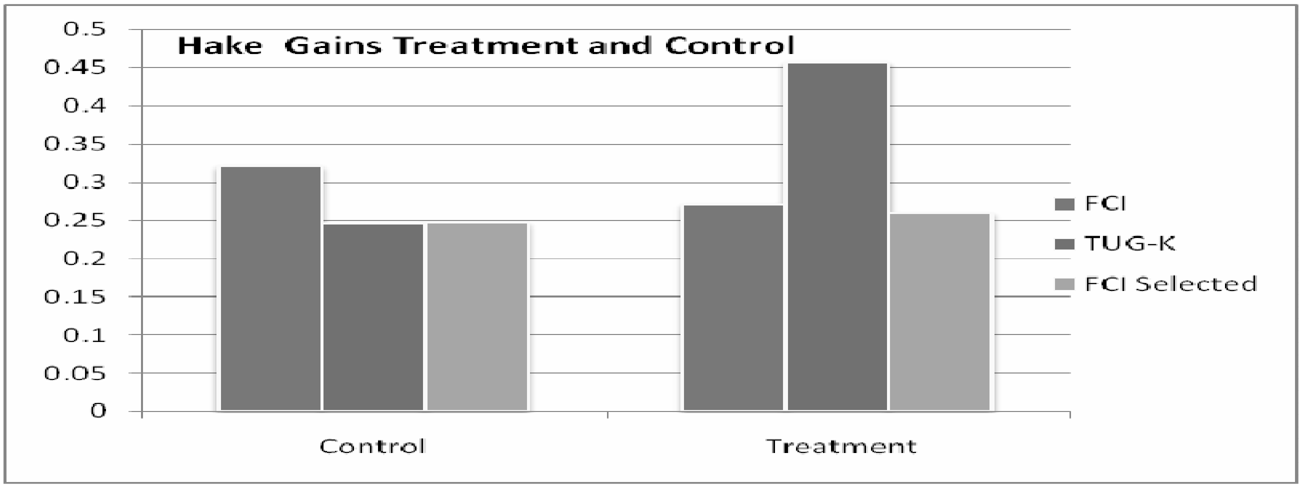


Figure 4. Histogram showing the average Hake gains, <g>, for the treatment and control groups on the FCI, TUG-K and FCI Selected questions.

I believe that because the average Hake gain for Archambault's group was more than 1.5 standard deviations above the average Hake gain of the control group, I believe that the treatment resulted in significantly greater gains in my students' kinematics understanding than the control group. Furthermore, Archambault's average Hake gain of 0.57 falls into the medium gain range for Hake scores while the control group's Hake gain of 0.24 falls into the low gain range for Hake scores. These results are consistent with the results of the treatment group overall and support our belief that our treatment significantly improved our students' understanding of kinematics concepts.

Qualitative Data

Upon reviewing the qualitative data for Archambault's group, it did not differ substantially from the data from the treatment group overall. However, it is worth noting one particular point that is consistent within the data from the treatment group. The graph below shows the accuracy of the responses from Archambault's students broken out by their method of solution.

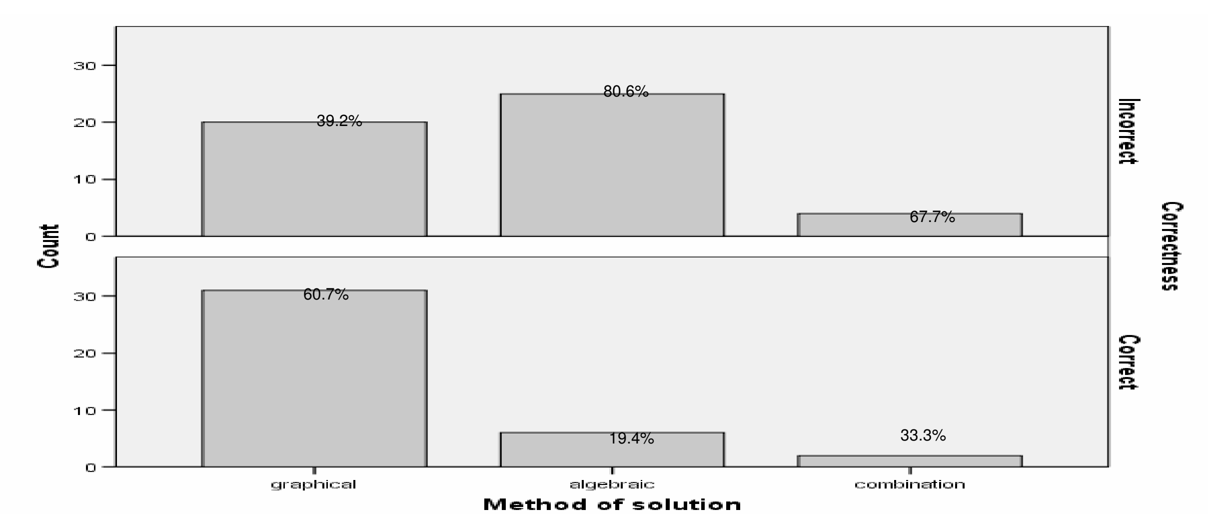


Figure 5. Histogram of Archambault's student responses to strip question 2 broken out by method of solution.

This clearly demonstrates that the students in Archambault's group were much more likely to be accurate with their solution if they solved by graphing exclusively (61 percent) than if they solved solely algebraically (19 percent) or in combination (33 percent). When the other investigator's groups are looked at individually, it is found that of the students that solved solely by graphing, about half of the students correctly solved the problem.

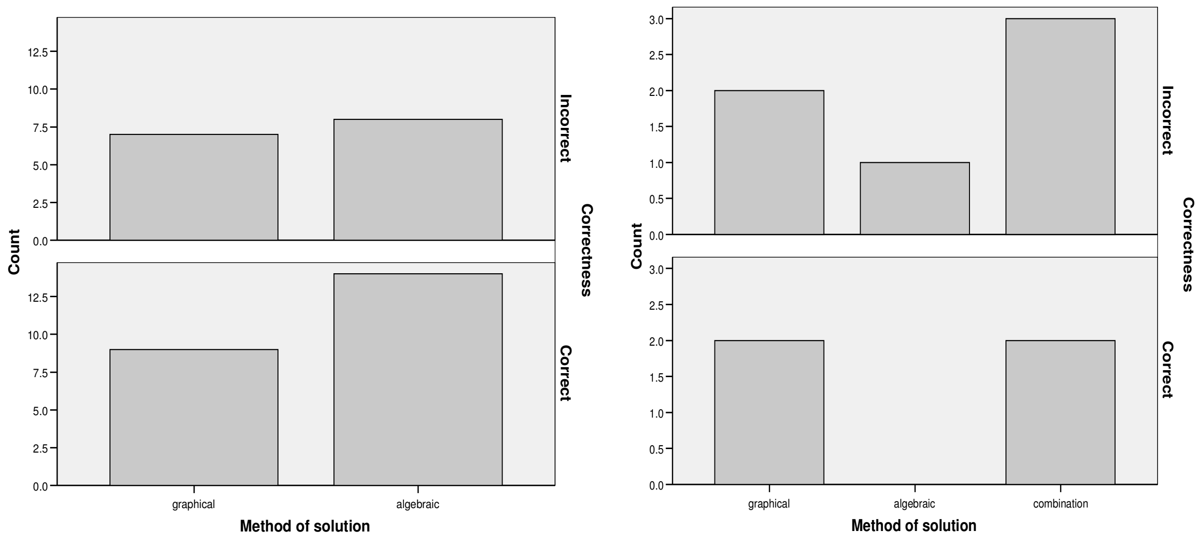


Figure 6. Histograms of investigator 3 and 4's student responses to strip question 2 broken out by method of solution. Investigator 3's students are on the left and investigator 4's students are on the right.

Even 3 of the 4 students in the control group who solved exclusively by graphing accurately solved the problem. It would certainly be appropriate to determine through further investigation if graphing is a better method for solving kinematics problems for high school students. The biggest problem appears to be convincing the students to use graphing to solve problem in the first place, but our treatment has been very successful in accomplishing that.

It is interesting to note, even if it is of very limited statistical significance, that we did a t-test comparing the posttest TUG-K scores from the Archambault group and the control group before we determined that the pretest groups were from different populations. We did this test because we were interested in determining whether the treatment samples and the control samples from the TUG-K posttest were from the same population. Based on a non-directional independent samples t-test at alpha = 0.05, we fail to reject the null hypothesis that population mean of the treatment group is equal to the population mean of the control group, $t(123) = 0.292$, $p = 0.771$. We conclude that there was not a significant difference between Archambault's group sample (Mean = 13.16, SD = 3.124) and the control sample (Mean = 13.35, SD = 3.676) on the TUG-K posttest. We are 95 % confident that the interval -1.071 to 1.441 contains the true population mean difference. (15 in Appendix A)

The interesting note is that the pretest TUG-K scores showed the students in the control group to be so much higher that they must have come from a different population than the Archambault group. But after the treatment, the two samples were from the same population. This information generally supports our results that we significantly improved our students' understanding of kinematics concepts.

Conclusion

I was very pleased that I was able to determine a strong positive effect from the treatment on my sample of students. As the semester progressed, it seemed that my students had a different understanding of kinematics than my students from previous years, but I was unable to tell without analyzing the data whether that difference was positive.

Even without the positive effect on students' understanding of kinematics, I think that I would probably continue to teach kinematics with a graphing first approach. There are advantages to be found throughout the year as we look at various graphical relationships among variables when the students are practiced in looking at these relationships.

It is also a fact that most of the high school physics students that we teach will not desire to complete science degrees, and the primary benefit that students get from taking high school physics is in problem solving skills and analyzing information generally. We would be remiss in not commenting on the practical benefits to students for their lives of being able to appropriately analyze a graph to determine what information is or is not present. I believe that our students benefitted significantly from this treatment in that regard.

Investigator 2 Field Trial Report

Introduction

Investigator 2 is Theresa Burch-Lococo from Phoenix, Arizona. This year was my 7th year of teaching physics. I taught at Cactus Shadows High School in Cave Creek, Arizona. There were 2 sections of first-year IB Physics (hereto forward referred to as "Honors Physics") and 2 sections of Conceptual Physics (hereto forward referred to as "General Physics") in my schedule this year. The combined total of students in all 4 sections was 90, with an approximately equal number of students in general physics as in honors physics (46 and 44 students respectively.)

At Cactus Shadows all students are required to take a physics course after biology. Some take chemistry first, then physics as juniors or seniors. Some take physics first, then chemistry. There were very few sophomore students in my classes; mostly juniors and seniors. This is partly because a passing grade in Algebra 1/2 is a pre-requisite to physics.

Procedure

As for the implementation of the treatment I followed very much the same procedures as my fellow investigators. Cactus Shadows offers six 50-minute classes every day, so all my physics classes met every day for 50 minutes. Because the classes were comprised of mostly juniors and seniors there were a variety of math levels present in the classes. The honors students were all in advanced math classes concurrently with physics. The majority of honors physics students were taking Precalculus. The students in the general physics course had all taken and passed Algebra 1/2 but many had passed with a "D" which really does not equate to a good working knowledge of algebra. Over half of my students still struggled with solving a simple algebra equation, while the other half could do it easily. These variances in math ability caused the class to move more slowly as it was necessary to interject algebra tutorials at various strategic points in the curriculum.

The pretests for the FCI and the TUG-K were given the first week of classes in August 2007 before any physics instruction was given. The posttests were given the last week of the first semester, in December 2007. Unfortunately, that last week of the semester many of my students were excused from class for various activities and missed the posttest. They all were required to take the test as make up work because it was figured in their first semester grade. However, many students realized that at that late date, the test grade had very little impact on their final grades. I feel that they did not take the posttests seriously for that reason. The evidence for my opinion about the posttest scores is that many of them fell below the pretest scores, which was not consistent with the other work exhibited by these students in the class. I feel that my FCI and TUG-K post scores are compromised by this portion of the population, but I was unable to find a criterion to exclude these scores in a way that would not compromise the integrity of the study.

Results

Tug-K

Initially I was interested in looking at my students' TUG-K scores to see if they had a significant gain in kinematics understanding over the course of the treatment. Based on a non-directional paired samples t-test at alpha = 0.05, I reject the null hypothesis that the population mean of the pretest takers is equal to the population mean of the posttest takers, t(90) = -12.963, p<0.001. I conclude that there was a significant difference in kinematics understanding as measured by the TUG-K for students before the treatment (Mean = 5.01, SD = 3.128) and after the treatment (Mean = 9.60, SD = 4.487). I am 95% confident that the interval -5.292 to -3.885 contains the true population mean difference. The correlation was 0.664. (16 in Appendix A)

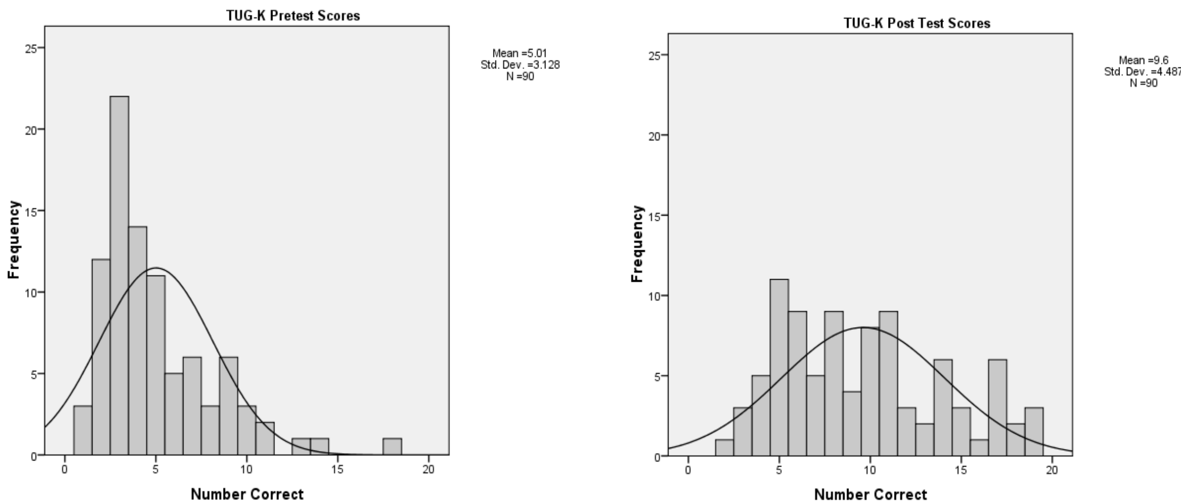


Figure 1. Comparison of Histograms for Burch-Lococo's TUG-K pretests and posttests.

The histograms in Figure 1 show that the pretest sample had a high frequency of low scores, and the posttest sample contained considerably more high scores. Because the posttest sample (Mean = 9.6) is shown to be from a population with a higher mean score (Mean = 5.01), I conclude that the students showed significant gains in their understanding of kinematics concepts after receiving the treatment. It seems that my students performed similarly to the treatment group overall, but I fear the aforementioned lack of concern for the posttest causes the effect to be less dramatic than the overall treatment group.

Next I wanted to know how my students compared to the control group so I set about to determining whether my sample and the control sample from the TUG-K pretest were from the same population. Based on a non-directional independent samples t-test at alpha = 0.05, I reject the null hypothesis that the population mean of the treatment group is equal to the population mean of the control group, t(84) = 7.144, p<0.001. I conclude that there was a significant difference between the treatment sample (Mean = 3.85, SD = 1.837) and the control sample (Mean = 9.40, SD = 4.893) on the TUG-K pretest. I am 95% confident that the interval 4.007 to 7.098 contains the true population mean difference. (17 in Appendix A)

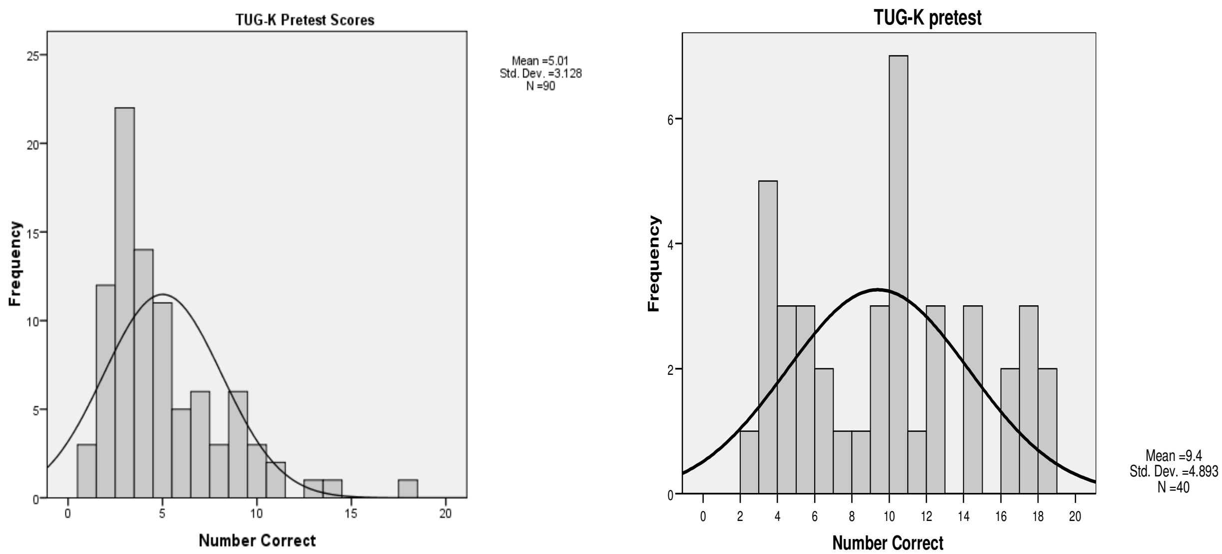


Figure 2. TUG-K pretest histogram on the left. Control TUG-K pretest histogram on the right.

Statistical testing shows that my group and the control group come from different populations. (17 in Appendix A) I conclude that it is inappropriate to compare TUG-K test scores between these two groups any further. My students have similar performance to the rest of the treatment group, but with lower posttest scores due to the aforementioned problem with the posttesting.

Because I was in possession of both pre and posttest data for my students, I was interested in the effect of the treatment on Purdue Spatial Inventory (PSI) scores. Particularly of interest was the relationship between gender and gains in PSI scores. Previous research states that "average scores on spatial tests for males are higher than those for female" and this "gender difference is often attributed to cultural factors (e.g. boys are more inclined than girls to play with construction toys, engage in motion-intensive sports, and play computer games)" (Hake, R.R., 2002). This article also refers to research studying increases in spatial ability, showed that women engineering students at Michigan Technological University could perform as well as men on spatial visualization tests if brought up to speed by a one-quarter (6hr/week) visualization course (Baartmans and Sorby, 1996).

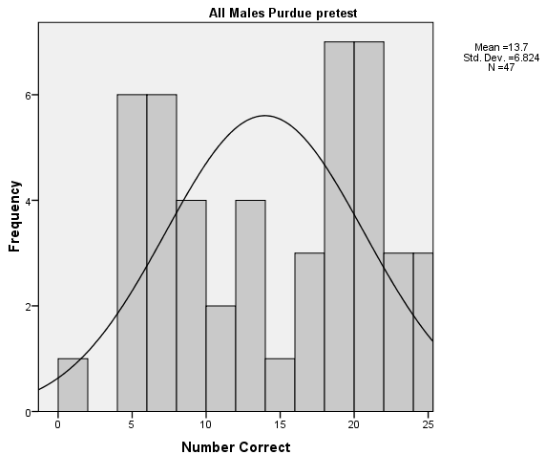


Figure 3a. Male PSI pretest

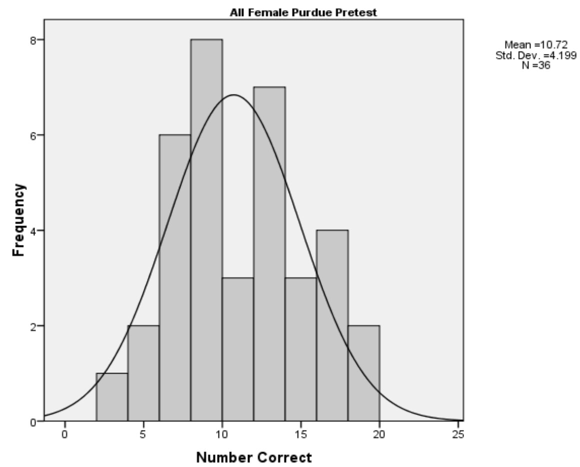


Figure 3b. Female PSI pretest

The histograms above (Figures 3a and 3b) depict PSI pretest scores for my male and female students. (18 in Appendix A) Clearly the male scores are higher, with a mean score of 13.70 correct answers. The female mean pretest score is 10.72 correct answers. To determine if this difference is statistically significant, a non-directional independent samples t-test at alpha = 0.05 was conducted. Based on this test the null hypothesis that the population mean of the male group was equal to the population mean of the female group was rejected, $t(81)=2.305$, $p=0.024$. I concluded that there was significant difference between the male sample (Mean = 13.70, SD = 6.824) and the female sample (Mean = 10.72, SD = 4.199) on the PSI pretest. I am 95% confident that the interval 0.408 to 5.552 contains the true population mean difference. Therefore, I assume that my students are average, with the males having higher average score on spatial tests than the females. The types of cultural factors that may have affected the groups were not studied other than the fact that all students were subjected to the graphical treatment.

Initially, I wanted to determine if, indeed, my students as a group had increased their PSI scores from the beginning to the end of the treatment. This was accomplished with a t-test and Hake gain analysis. Then I compared the pre and post test gains (or losses) across different groups. Finally, I analyzed the PSI score gains for male and female groups specifically.

The next test compared the PSI pretest scores of all of my students to their posttest scores. This would determine if the students had a statistically significant difference in scores. Based on a non-directional dependent sample t-test at $\alpha = .05$, I reject the null hypothesis that the population means are equal, $t(82)= 4.023$, $p< 0.001$. It is concluded that there is a significant difference in Purdue Spatial Inventory Pretest scores (Mean = 12.41, SD = 5.99) and Purdue Spatial Inventory Posttest scores (Mean = 14.11, SD = 5.52). There is 95% confidence that the interval 2.5 to 0.86 contains the true population mean difference. The correlation was 0.78.

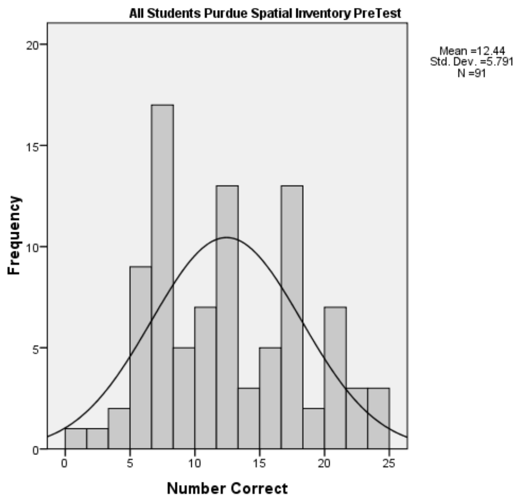


Figure 4a. All PSI Pretest

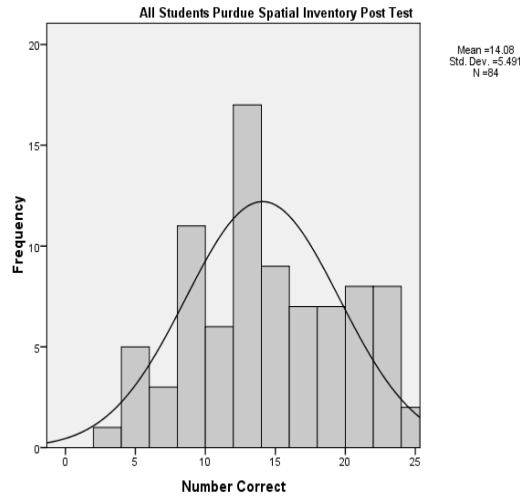


Figure 4b. All PSI Posttest

This test (Figure 4a and 4b) reveals that the observed difference between pre and posttest scores is most likely a statistically significant one. This means that when the students took the posttest, they were somehow a different population than they were when they took the pretest, a population with significantly more spatial ability. It is possible that other factors in the students' environment could have caused the rise in scores (such as certain math courses), but since they all received the treatment we assume that it is a likely candidate to have had some influence.

Pre and posttest scores were averaged. An average pretest score of 12.40 correct answers was found for all my students together. The average posttest score for the same group was 14.11 correct answers. The average Hake gain for pre and post test scores was calculated to be 0.147. (19 in Appendix A)

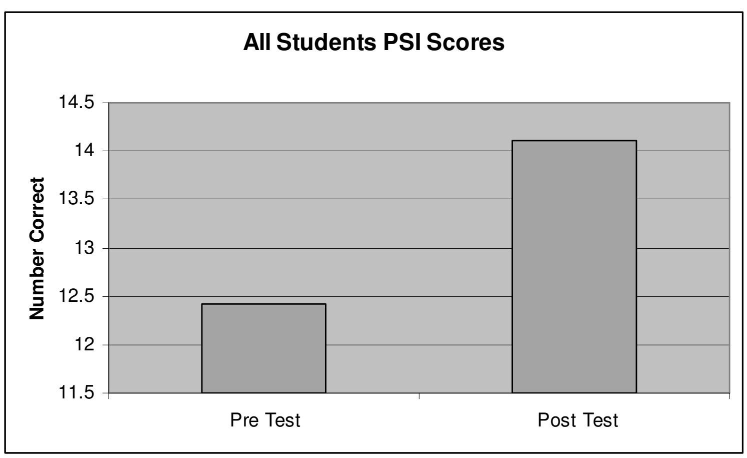


Figure 5. All Students PSI Pre and Posttest

Figure 5 shows an overall positive gain for all of my students from pre to posttest score. The students were further broken out by gender alone, level alone, and gender/level combined, test scores averaged and Hake score calculated and averaged. Figure 6 below shows the pretest scores in the foreground for each group. Posttest scores for the corresponding groups are shown behind the pretest scores. All groups show increases from pre to post test.

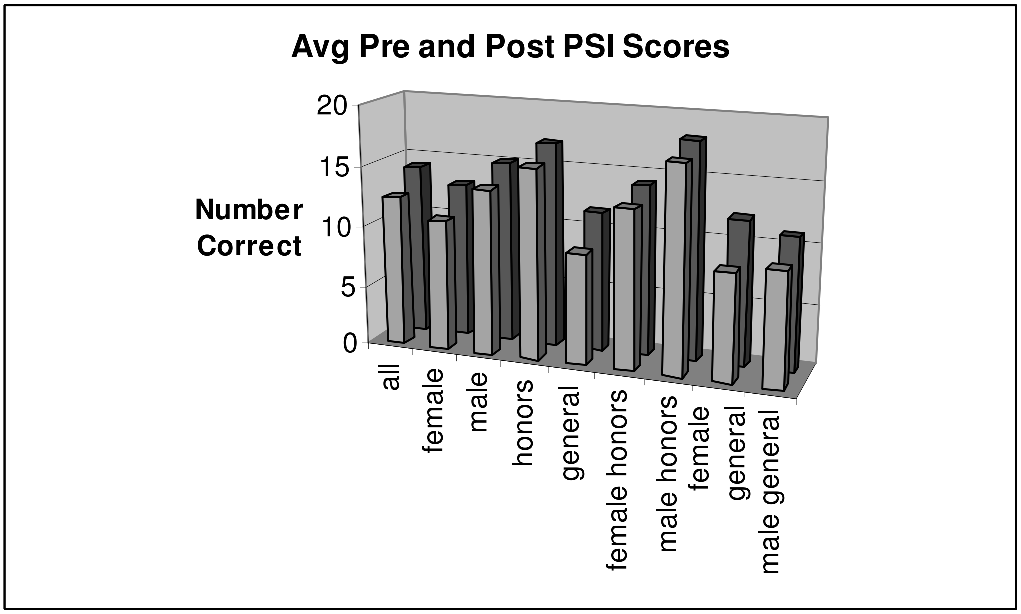


Figure 6. Average Pre and Post PSI Scores

Figure 6 above shows that all groups exhibit an increase in PSI test scores from the beginning to the end of the treatment. Hake scores were calculated to analyze these increases.

Group Number	Group	Purdue Pretest Average Score	Purdue Posttest Average Score	Hake gain Average
1	All	12.41	14.11	0.147
2	Female	10.77	12.86	0.158
3	Male	13.70	15.02	0.128
4	Honors	15.79	16.95	0.142
5	General	9.24	11.55	0.156
6	Honors Female	13.2	14.13	0.086
7	Honors Male	17.08	18.04	0.138
8	General Female	8.95	12.04	0.206
9	General Male	9.52	11.05	0.105

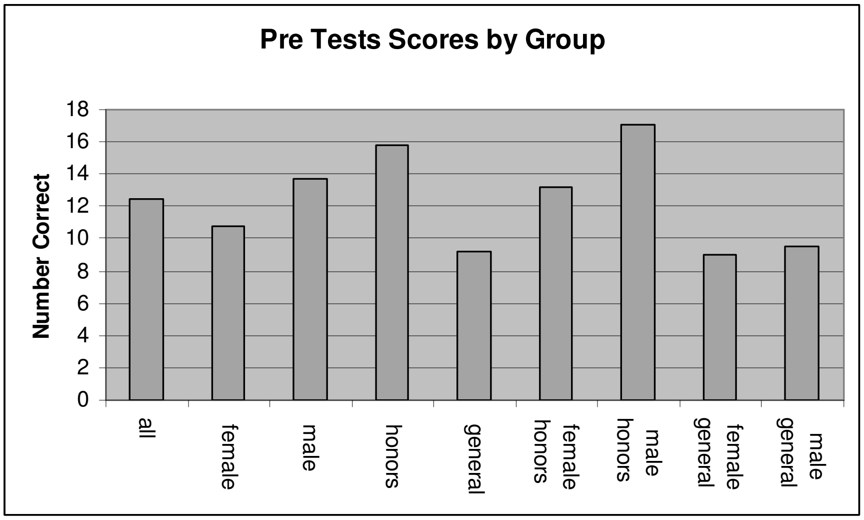


Figure 7. PSI Pretest Score by Group

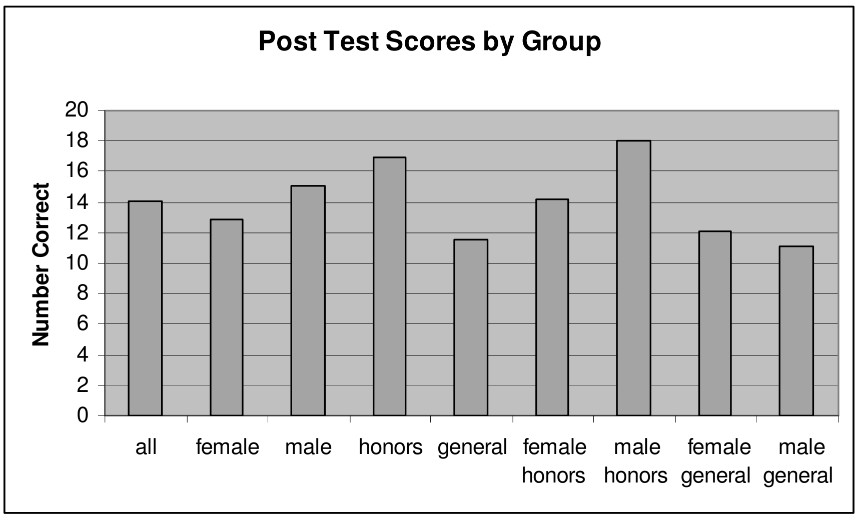


Figure 8. PSI Posttest Score by Group

Figure 7 identifies the group with the highest pretest score as honors males. The lowest pretest score is general females. As shown in Figure 8, honors males remain highest in posttest scores but general males are now lowest, although all scores reflect an increase.

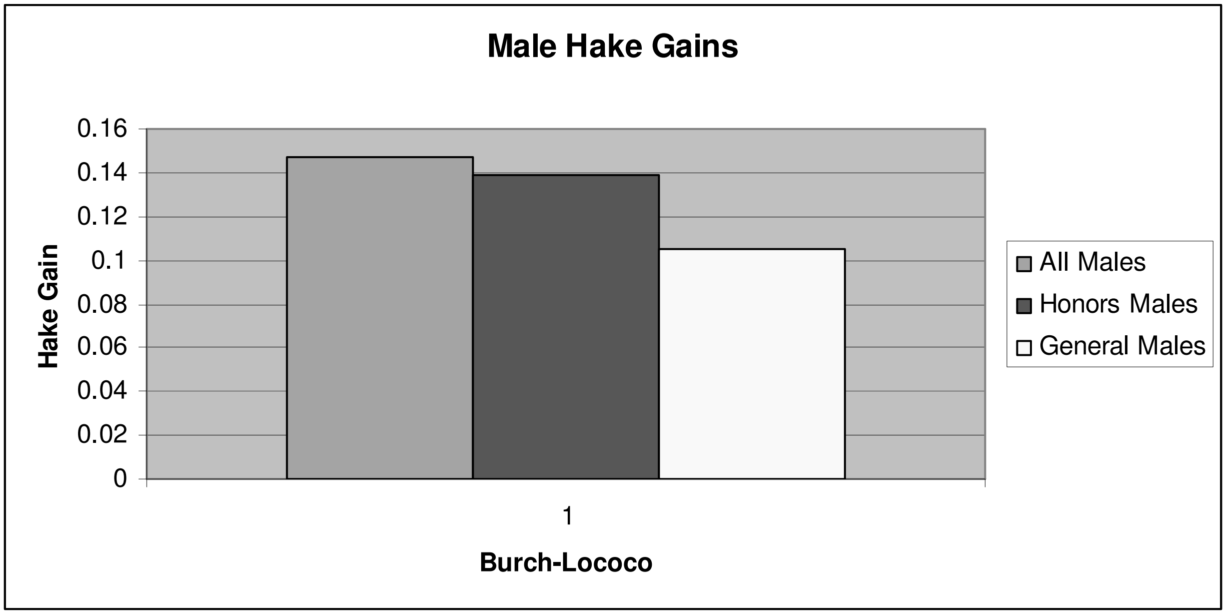


Figure 9. PSI Male Hake Gains

The Hake gain measures the ratio of actual improvement to possible improvement. In Figure 9 honors males show a Hake gain of 0.139 from pre to posttest and general males show a Hake gain of 0.105. This means that even though the honors males came in with more spatial ability, they actually improved more than the general males. The honor males exhibited a higher proportion of their possible improvement than the general males. But, all male participants did show an improvement in spatial performance.

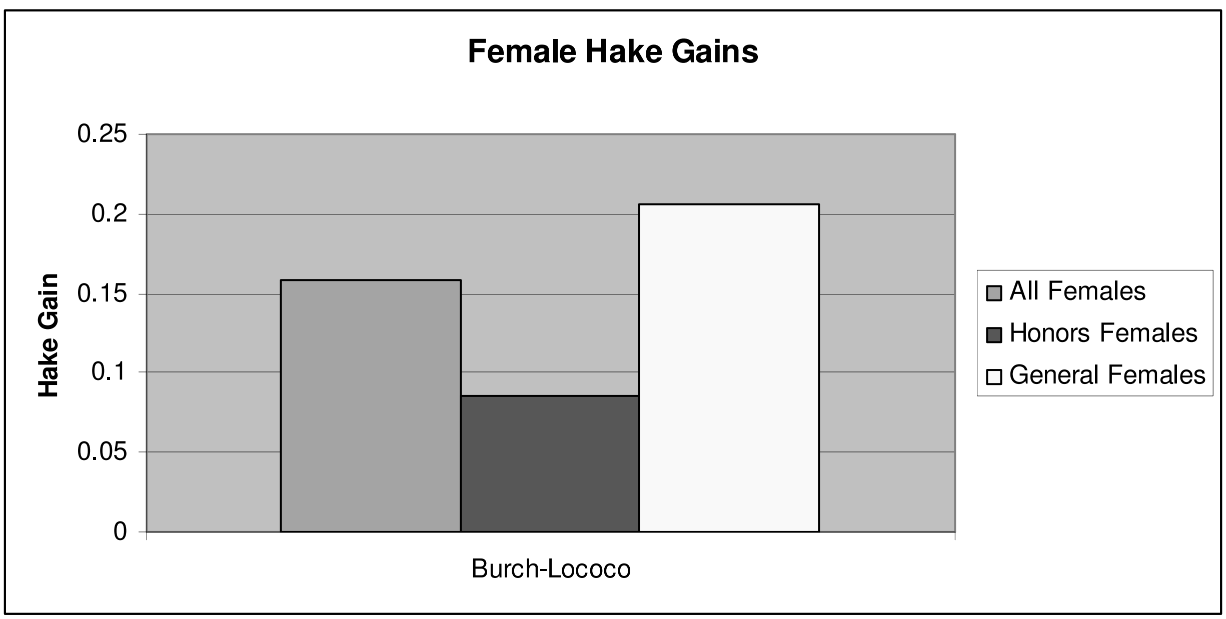


Figure 10. PSI Female Hake Gains

Similarly, all female participants improved in spatial performance. In Figure 10 honors females show a Hake gain of 0.086. This is the lowest Hake gain of any group in the study. The next lowest Hake gain was that of general males. General females, however, exhibited a Hake gain of 0.205, which is the highest Hake gain in the study. This means that even though the general females came in with very low spatial scores, they improved more than any other group in the study. This is consistent with the Baartman and Sorby study cited earlier, stating that female spatial scores can be brought up to speed with a visualization course. If the treatment is taken to be a contributing factor to this improvement, then it follows that the graphical treatment is comparable to a visualization course.

Why, then, did honors female participants not benefit from the treatment as much as the general females? It could be that it is not the treatment at all causing the spatial differences but some other factors like math courses. Honors males and female take approximately the same courses, yet exhibit different Hake gains for the PSI. Perhaps there is a sort of natural selection occurring due to the fact that many female students who do not have high spatial ability in the first place tend not to register for advanced/honors physics. This leaves those who have already raised their spatial abilities before entering the class, therefore leaving less room for improvement. At Investigator 2's high school, all students are required to take physics, so female students entering the general class are more likely to enter at levels resulting from the cultural factors referred to in the opening paragraphs. They are more likely to experience all improvement in spatial ability during the course of the treatment. According to previously cited research the lower Hake gains for honors and general males are consistent with previous findings as well.

Next, I wanted to compare PSI scores to TUG-K scores. A correlation was run between PSI pretest scores and TUG-K pretest scores for all of my students.

Correlations

		Purdue Spatial pretest	TUG-K pretest
Purdue Spatial pretest	Pearson Correlation	1.000	.490**
	Sig. (2-tailed)		.000
	N	88	88
TUG-K pretest	Pearson Correlation	.490**	1.000
	Sig. (2-tailed)	.000	
	N	88	88

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

Figure 11. PSI/TUG-K Pretest Correlations

The correlation between the PSI pretest and the TUG-K pretest (see Figure 9) was moderate, 0.490. This is slightly higher than the same test for all treatment students (corr = 0.345). This means that students who do well on the PSI pretest tend to also do well on the TUG-K pretest, and students who do not do well on the PSI pretest tend to do poorly on the TUG-K pretest as well.

Next, a correlation was run between PSI pretest scores and TUG-K posttest scores for the same sample.

Correlations

		Purdue Spatial pretest	TUG-K post test
Purdue Spatial pretest	Pearson Correlation	1.000	.598**
	Sig. (2-tailed)		.000
	N	88	88
TUG-K post test	Pearson Correlation	.598**	1.000
	Sig. (2-tailed)	.000	
	N	88	88

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

Figure 12. PSI Pre/TUG-K Posttest Correlations

The correlation between the PSI pretest and the TUG-K posttest (see Figure 10) was also moderate, 0.598. This is slightly higher than the same test for all treatment students (corr = 0.398). This means that students who do well on the PSI pretest tend to also do well on the TUG-K posttest, and students who do not do well on the PSI pretest tend to do poorly on the TUG-K posttest as well.

Finally, a correlation was run for this sample comparing PSI posttest scores and TUG-K posttest scores for all of Investigator 2's students.

Correlations

		Purdue Spatial post test	TUG-K post test
Purdue Spatial post test	Pearson Correlation	1.000	.581**
	Sig. (2-tailed)		.000
	N	82	82
TUG-K post test	Pearson Correlation	.581**	1.000
	Sig. (2-tailed)	.000	
	N	82	82

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

Figure 13. PSI Post/TUG-K Posttest Correlations

The correlation between the PSI posttest and the TUG-K posttest (see Figure 11) was again moderate at 0.581. There is no PSI posttest data for the overall treatment sample, so no comparison can be made. The correlation means that students tend to do as well on the PSI post test as they do on the TUG-K posttest. The correlation tests in general show that students' spatial thinking skills as measured by the PSI did tend to play a role in performance on the TUG-K for my smaller student sample.

I wanted to compare the Hake gains on the PSI to the Hake gains on the TUG-K for students grouped by level and gender.

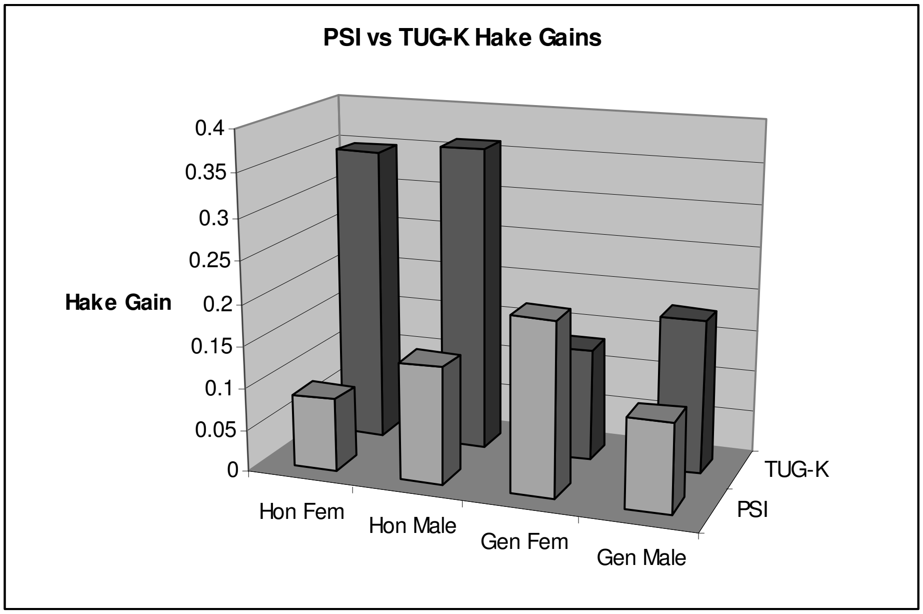


Figure 14. PSI vs TUG-K Hake Gains

Figure 14 above compares the average PSI Hake gains to the average TUG-K Hake gains. While all groups gained on both tests, the greatest Hake gains appear to be on the TUG-K tests. Of interest are the TUG-K Hake gains for the honors students. Although the PSI Hake gains were significantly lower for honors females (0.086 females vs. 0.139 males) the TUG-K Hake gains of honors males and females are almost identical (0.350 vs. 0.364). The TUG-K Hake gain of the honors females is approximately 4 times that of their PSI Hake gain, while honors males have a TUG-K Hake gain that is only about 2 1/2 times that of the PSI Hake gain. This seems to be consistent with the research that female students can "catch up" to male in spatial ability.

In conclusion, the low PSI Hake gain still translated to a sizable TUG-K Hake gain for the general males group, although not as dramatic as that for the honors group. The TUG-K Hake gain of 0.182 is approximately 1.7 times the gain exhibited for the PSI. General females show the most interesting result of all. Although they had the highest Hake gain of all groups on the PSI, the general females group had the lowest Hake gain on the TUG-K. The TUG-K score was only slightly more than half the gain on the PSI. It seems that although they are correlated, spatial skills do not necessarily mean students will do better at solving problems graphically. So apparently there is more to it than spatial ability. There is evidence that spatial thinkers do better at problem solving, but I am seeing students who have large gains in spatial ability and small gains in problem solving skills. Overall the treatment seems to be a contributing factor in helping students to become better problem solvers and increasing spatial ability, but these two abilities are not necessarily linked to each other.

Student Problem Solving

Although students who use graphing seem to be better problem solvers, I believe that this effect may be even more pronounced than the data actually shows. Given the strip problem data that was discussed in the body of this report, it was shown that in the treatment group, students who used a graphical approach solved problems correctly more often than students that solved algebraically. Students in my classes were asked to solve problems as they explained aloud what they were doing. These videotaped interviews reveal that even when a student's paper shows only algebraic solutions, often the algebra is based on a mental graphical picture that the student is drawing from. A student may write down the algebraic formula, but he knows to use it because he is relating the solution to, say, the area under the curve of a graph. Such connections made mentally do not show up on a student's paper and may be mislabeled as algebraic solutions instead of graphical or combination. I suspect that more of the treatment students were actually graphical problem solvers than originally thought.

A videotaped interview with "Student A" is representative of this phenomenon. Student A was given a problem to solve. Originally, when he solved the problem in class, he used a combination approach with a partial algebraic and a partial graphical solution. When asked to explain the steps he used to solve the problem, Student A related a procedure that really should have been categorized a graphical, not combination. The following transcription of the solution to the "Ella Vader" strip problem will help to illustrate this.

Student A: All right…so in this problem you know that the girl is driving a Porsche on a highway at 30 m/s … and so … you know that she continues on for that speed for 2 s … so that's 30 times 2 to find your distance traveled … **(Student writes out a line of algebra to represent this calculation)** … that's 60 … and then you know that she decelerates when she sees a deer after 7.5 s and I guess we just assume that it's a constant deceleration … **(students begins drawing a graph)** …so you have your speed right here … 30 m/s … time is … here on the graph you have 7.5 s down here and … this goes like that … and all this under here is the distance she traveled … total distance … and so the way you find that is you just find the area of this triangle so that's 30 times 7.5 over 2 which equals 112.5 meters and so to find the total distance she traveled in the 9.5 seconds you just add 60 plus 112.5 and that gives you 172.5 meters which is the answer.

I contend that although Student A solved the first part of the problem with algebra, it was in response to the mental picture of a rectangular area under the curve of a velocity vs. time graph. This seems evident because later when the more complicated shape (the triangle) came along, Student A sketched the graph because it was slightly more difficult to hold the image in his head. By drawing the graph he was able to keep himself organized as the problem became more involved. It seems obvious that the student was using a mental picture of a velocity vs. time graph to solve the problem from the beginning.

Algebraic representations of simpler parts of problems with graphing coming into play as the problem progresses were common in our study. This seems to suggest that graphical solutions have become part of the students' thought process to the point that they have become automatic. It follows that the students with higher spatial ability would be better able to hold these graphical images in their heads and therefore be less likely to draw them out as part of the solution to a problem. Perhaps there are more graphical problem solvers in the study than originally thought. This would cause the apparent overwhelming preference for algebraic solutions to be overstated and the graphical correctly solved problems to be understated.

Investigator 3 Field Trial Report

Procedure

My name is Michael Crofton and I teach physics in a suburb of Minneapolis, Minnesota. Last year all of my 38 honors physics students were part of this study. Of my 52 general physics students only 28 were part of the study. My school does not lock the students into the same teacher all year. Therefore some students had both the treatment teacher and the control teacher during the study. These students were all dropped from the statistics when the data was analyzed.

I did not administer the Purdue spatial test. This was due to the fact that I had lost a day to a late school start and had already taken two days for the other pretests that were part of the study. I did not want to lose another class day at the start of the trimester. For the study I administered the FCI pretest on the first day of school and the FCI posttest during the last week of school. The test was counted as part of the final exam. At the time they took the posttest my students had been done with mechanics for 10 to 12 weeks depending on which physics class they were in.

It is my belief that the most important part of the treatment was incorporating some new worksheets created to require the students to solve for the answers graphically even if the students were able to solve algebraically. Then in whiteboard sessions the students sketched a graph, used the sketch in their solution and explained their solution to their classmates.

Results for TUG-K

Honors Physics

The honors physics TUG-K pretest mean was 9.32 and the posttest mean was 17.63 out of 20.

Class	Test	Pretest mean	Posttest mean	Normalized Gain for pre and posttests
Honors Physics	TUG-K	9.32	17.63	0.80

There was no control group in honors physics with which to run a comparison therefore I compared the pre and posttests by running the Hake test for normalizing the gain and found it to be 0.80.

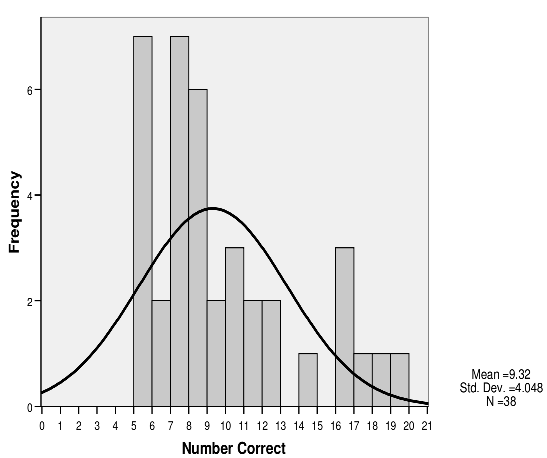


Figure 1. Honors TUG-K pretest

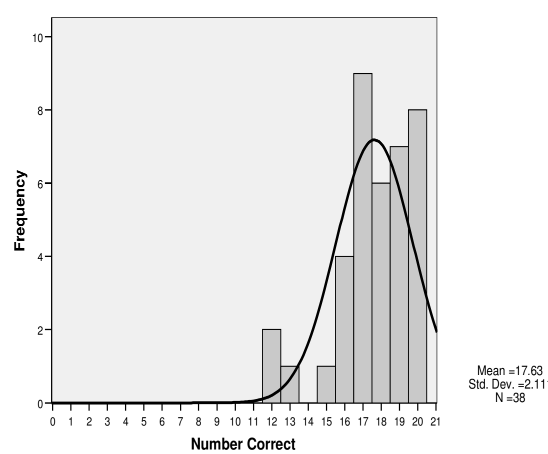


Figure 2. Honors TUG-K posttest

A directional dependent sample t-test was conducted to evaluate the hypothesis that the students in my honors physics class would increase their TUG-K score. At $\alpha = 0.01$ I reject the null hypothesis that the population mean for the posttest is equal to or less than the population mean of the pretest $t(37)=8.316$, $p<0.001$. I conclude that there is a significant difference in the scores of the students for the pretest (Mean=9.32, SD=4.05) and the scores on the posttest (Mean=17.63, SD=2.11). The results of the t-test indicate that the treatment my students received led to much higher scores on the posttest. There was a moderate correlation between the students scores of 0.50 meaning that many of the students that had high scores on the pretest had high scores on the posttest. (20 in Appendix A)

The pretest and posttest performance on the TUG-K was also analysed using the Hake test. It was determined that the average normalized gain for the group was $\langle g \rangle_{38}=0.80\pm0.17$. (21 in Appendix A) In the language of the creator of the Hake test the honors students have demonstrated very high gains. In studies that Richard Hake has done involving thousands of students he has found that gains in traditional instruction rarely exceed 0.30. (Hake, 1998)

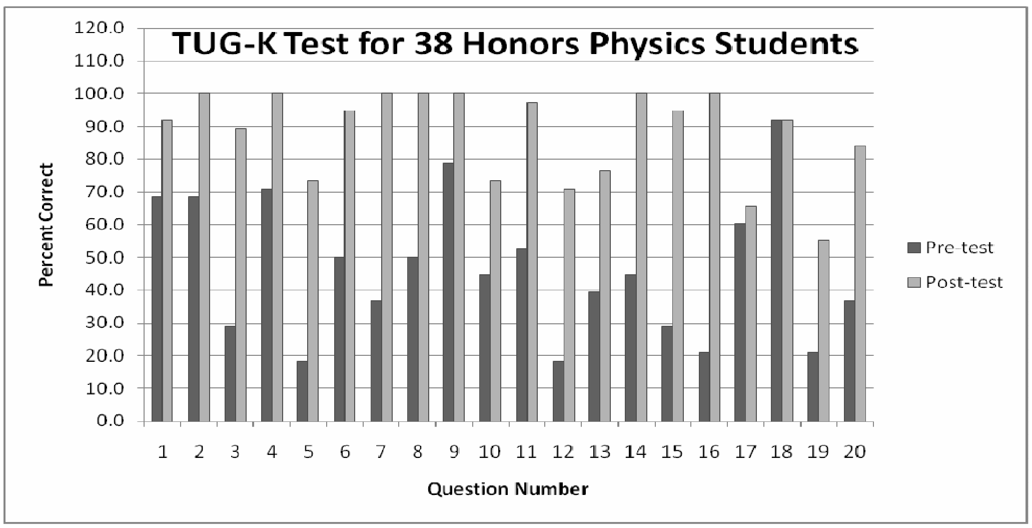


Figure 3. Honors physics performance on the 20 questions in the TUG-K test

I thought it would be interesting to look at whether or not the skills emphasized in the treatment, taking slopes and areas to solve problems, would yield mastery of those areas of the test. The questions that required students to take slopes and areas to solve for a value in the TUG-K are questions 3, 4, 5, 12, 13, 14 and 16. Questions 5 and 13 required them to calculate a slope to answer the question. However the two questions were a bit tricky as the line segment was in the middle of a graph and some students solved for slope of a line beginning at the y-axis. Most of the students that answered incorrectly knew they should take a slope to answer the question; however they calculated the slope incorrectly. Students were very comfortable in solving questions that required them to find the area of a graph except in question 12. That question required students to find the area of an acceleration vs. time graph to solve for the change in velocity. This question showed that the students were not as comfortable with the area of an acceleration vs. time graph as with a velocity vs. time graph. Overall the honors students demonstrated proficiency at close to the 90% level.

General Physics

The treatment general physics TUG-K pretest mean was 5.39 and the posttest mean was 13.68 out of 20 with a normalized gain of 0.57. The control group of general physics students had a pretest mean of 9.40 and a posttest mean of 13.35 with a normalized gain of 0.24.

Class	Test	Pretest mean	Posttest mean	Normalized Gain for pre and posttests
Treatment General Physics	TUG-K	5.39	13.68	0.57
Control General Physics	TUG-K	9.40	13.35	0.24

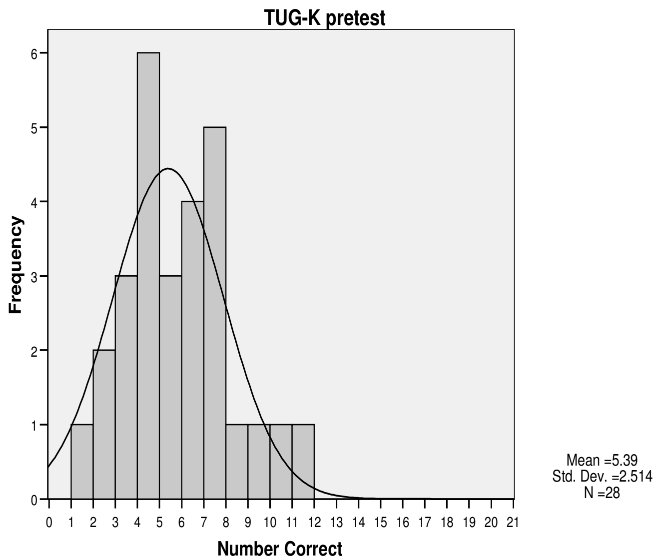


Figure 4. Treatment TUG-K pretest

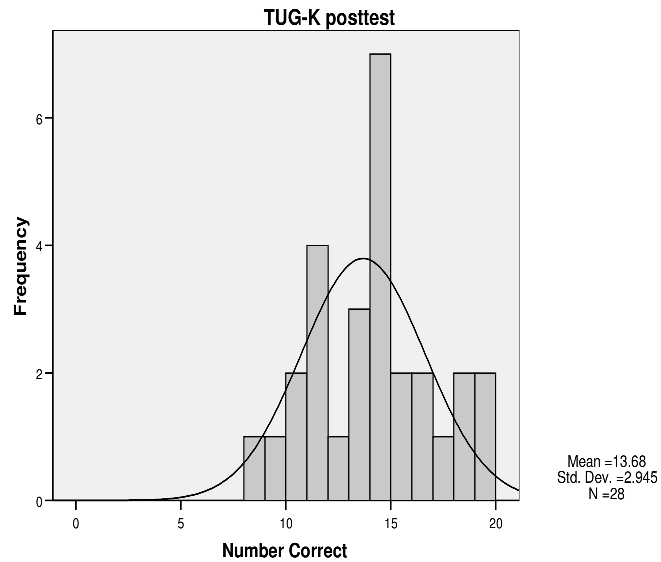


Figure 5. Treatment TUG-K posttest

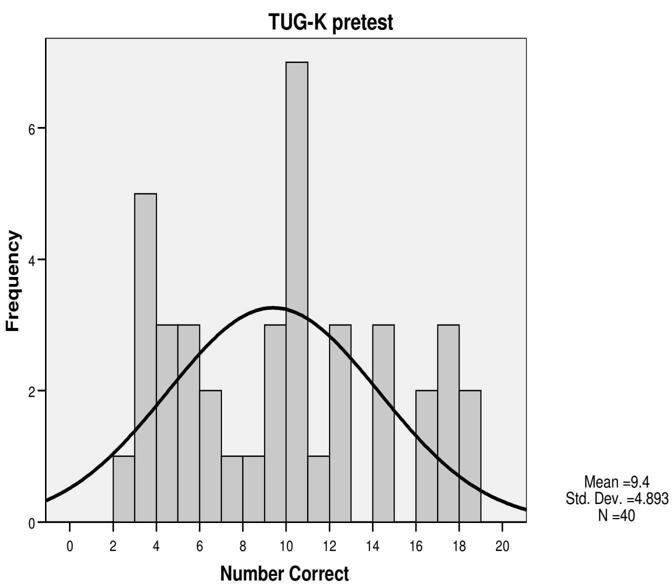


Figure 6. Control TUG-K pretest

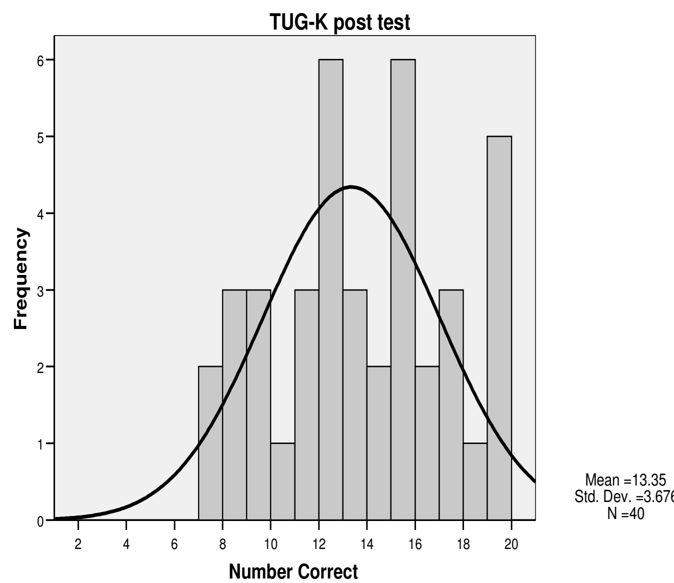


Figure 7. Control TUG-K posttest

I was interested in the effect of my graphical treatment on the students' TUG-K scores. Based on a non-directional independent samples t-test at $\alpha=0.05$, I fail to reject the null hypothesis that the population means are equal, t(66) = -0.393, p = 0.696. I therefore conclude that there is no significant difference between my treatment students (Mean = 13.68, SD = 2.95) and the control group (Mean = 13.35, SD = 3.67) for the TUG-K posttest. I am 95% confident that the true population mean difference falls between -2.00 and 1.34. This t-test showed that there was no difference between the scores of my students and those of the control students on the posttest. (22 in Appendix A)

An analysis was also run on the two populations for the pretest using a non-directional independent samples t-test at $\alpha=0.05$. In this case the null hypothesis that the two populations have the same mean was rejected, t(66) = 3.98, p <0.001. The conclusion is that the treatment group (Mean = 5.39, SD = 2.51) was a far different population than the control group (Mean = 9.40, SD = 4.89) at the time of the pretest. I am 95% confident that the interval 2.19 to 5.82 contains the true population mean difference. This t-test showed that the scores of the treatment students were much lower on the pretest than the scores for the control students. The means are so different that a meaningful study on the effects of the treatment using the t-test as an analysis tool is not valid. (22 in Appendix A)

The TUG-K performance of the treatment group and the control group was compared using the Hake test. Based on the average of the Hake gains of the students I determined there was a substantially larger gain in the treatment group $\langle g \rangle_{28T}=0.57\pm0.19$ than for the control group $\langle g \rangle_{28C}=0.24\pm0.51$. The treatment group mean gain was 1.74 standard deviations above the controls' mean gain. The Hake test shows that my treatment students had much greater gain on their scores than the control students. This is shown also by the fact that my treatment students scored much lower on the pretest but ended up with the same average. The gain of the treatment group showed medium gains while the control group showed low gains. (23 in Appendix A)

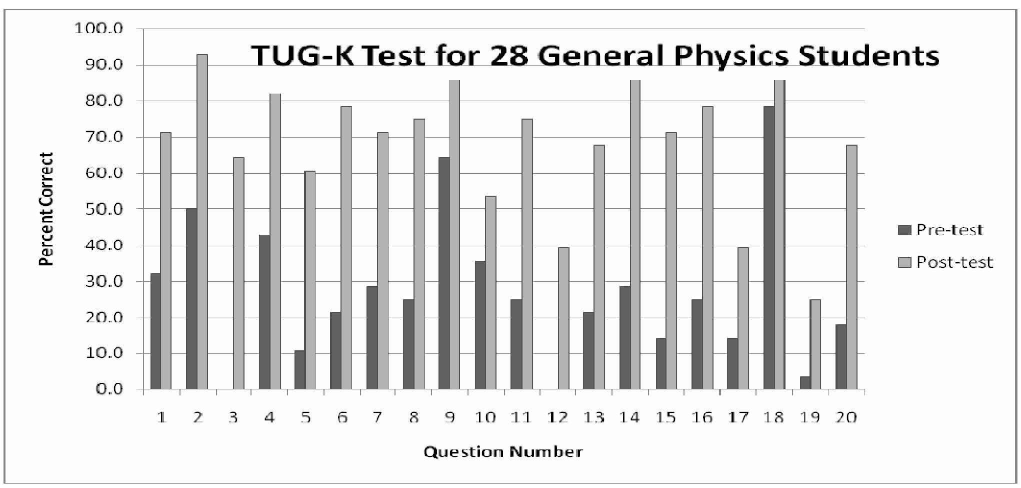


Figure 8. General physics treatment group performance on the 20 questions in the TUG-K test

Once again, the questions the treatment focused on are 3, 4, 5, 12, 13, 14 and 16. Two of the questions that my students did poorly on were addressed in the graphical treatment. Question 12 required the students to take the area of an acceleration vs. time graph to solve for the change in velocity. Obviously this skill wasn't fully understood by the students. Question 5 required taking the slope of a velocity vs. time graph to solve for the acceleration, a skill that was heavily emphasized. However, this graph does not begin at zero and many of the students did not notice this distinction. While most students did recognize they needed to take a slope, they did not do so correctly.

Results for the FCI

Honors Physics

The honors physics FCI pretest mean was 8.24 and the posttest mean was 23.71 out of 30.

Class	Test	Pretest mean	Posttest mean	Normalized Gain for pre and posttests
Honors Physics	FCI	8.24	23.71	0.72

There was no control group in honors physics with which to run a comparison therefore I compared the pretests and posttests by running the Hake test for normalizing the gain and found it to be 0.72.

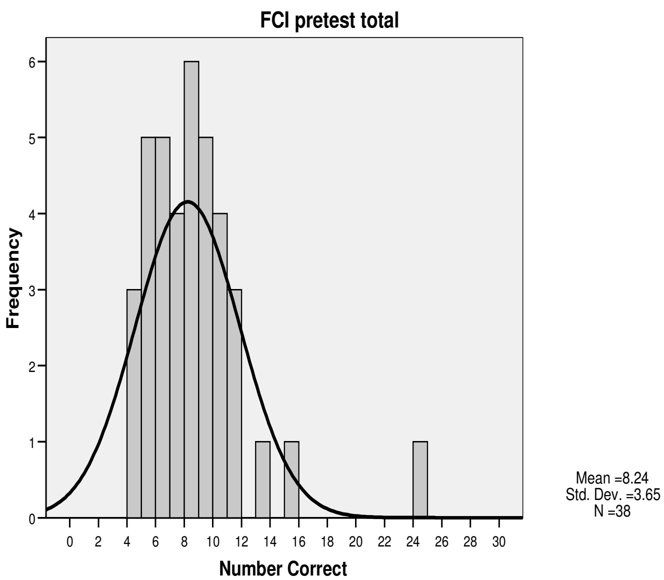


Figure 9. Honors FCI pretest

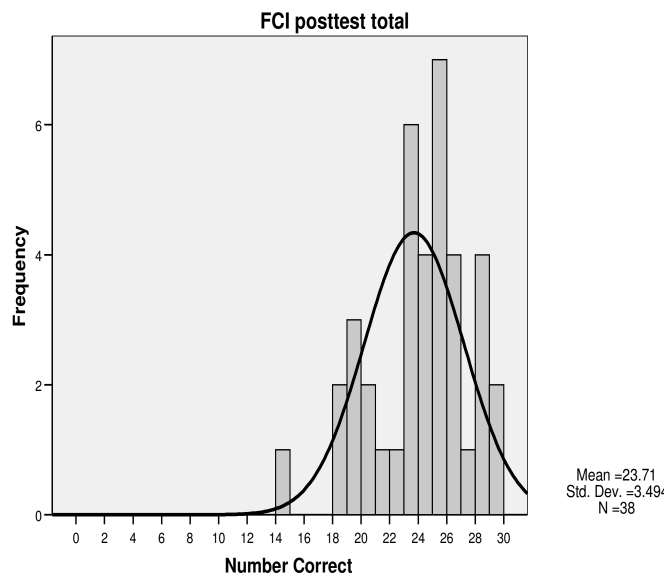


Figure 10. Honors FCI posttest

A directional dependent sample t-test was conducted to evaluate the hypothesis that the students in my honors physics class would increase their FCI score. At $\alpha = 0.01$ I reject the null hypothesis that the population mean for the posttest is less than or equal to the population mean of the pretest $t(37)=25.11$, $p<0.001$. I conclude that there is a significant difference in the scores of the students for the pretest (Mean=8.24, SD=3.65) and the scores on the posttest (Mean=23.71, SD=3.49). The results of the t-test indicate that the treatment the honors students received led to higher scores on the posttest. There was a moderate correlation between the students scores of 0.46 meaning that many of the students that had high scores on the pretest had high scores on the posttest. (24 in Appendix A)

The pretest and posttest performance of the honors students on the FCI was also compared using the Hake test. It was determined that the average gain for the group was $\langle g \rangle_{38}=0.72\pm0.15$. The honors students demonstrated high gains on the FCI test. (21 in Appendix A)

General Physics

The general physics FCI pretest mean was 7.39 and the posttest mean was 18.68 out of 30 with a normalized gain of 0.50. The control group of general physics students had a pretest mean of 9.40 and a posttest mean of 13.35 with a normalized gain of 0.24.

Class	Test	Pretest mean	Posttest mean	Normalized Gain for pre and posttests
Treatment General Physics	FCI	7.39	18.68	0.50
Control General Physics	FCI	8.19	15.14	0.32

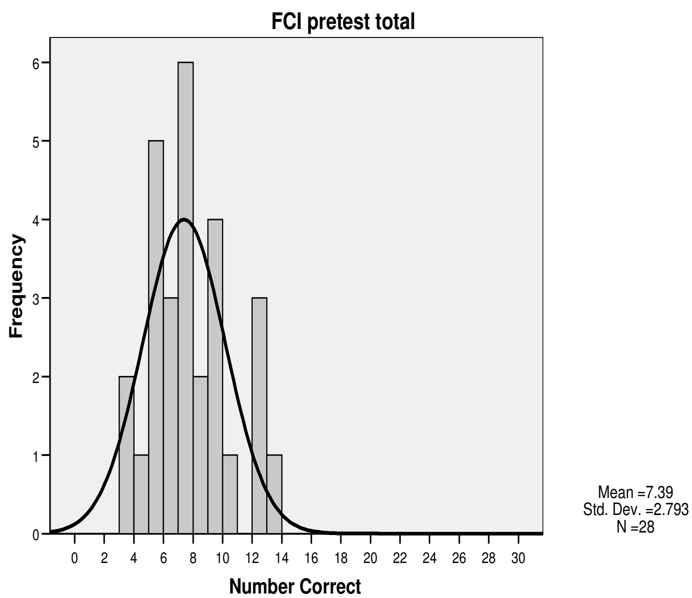


Figure 11. Treatment FCI pretest

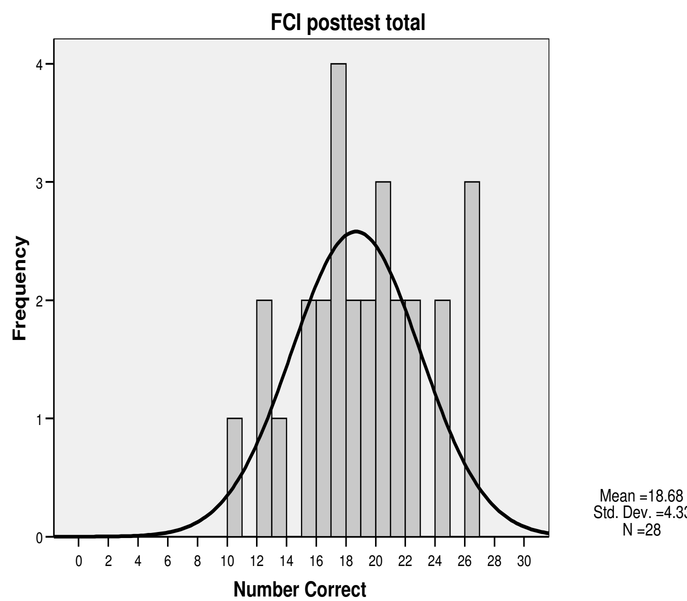


Figure 12. Treatment FCI posttest

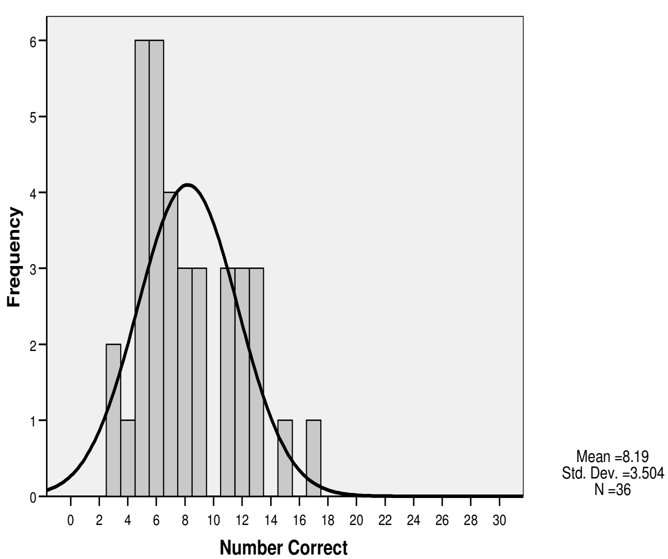


Figure 13. Control FCI pretest

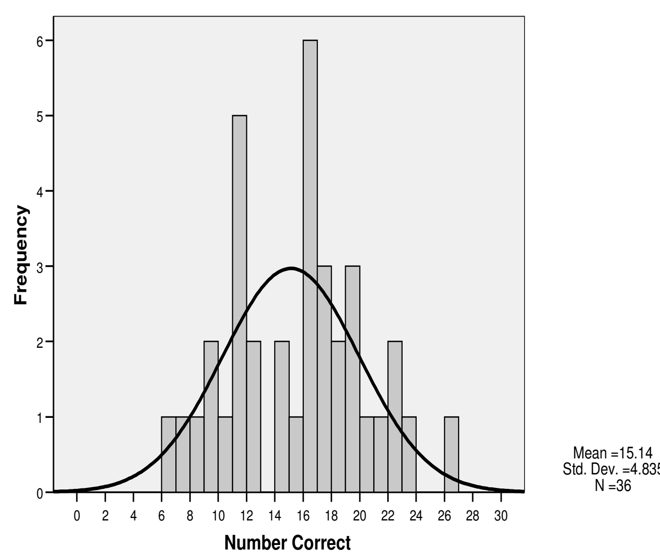


Figure 14. Control FCI posttest

I was interested in the effect of my graphical treatment on the students' FCI scores. Based on a directional independent samples t-test at α=0.01, I reject the null hypothesis that the population of mean treatment students is less than or equal to that of the control group, t(62) =3.04, p = 0.0015. I therefore conclude that treatment group has significantly higher scores (Mean=18.68, SD=4.33) than the control group (Mean = 15.14, SD = 4.84). (25 in Appendix A) I believe the results indicate that my general physics students had a much better understanding of forces and motion than the control group.

An analysis was also run on the two populations for the pretest. A non-directional independent samples t-test at α=0.01 was run. In this case I fail to reject the null hypothesis that the two populations have the same means, t(62) =0.99, p=0.326. The conclusion is that there is no significant difference between the treatment group (Mean = 7.39, SD = 2.79) and the control group (Mean = 8.19, SD = 3.50) at the time of the pretest. (26 in Appendix A) It is interesting to note that even though the populations were far different at the beginning of the year on the TUG-K test they were very similar on the FCI.

The FCI performance of the treatment group and the control group was compared using the Hake test. Based on the average of the Hake gains of the students I determined there was a substantially larger gain in the scores of the treatment group $<g>_{28T}=0.50\pm0.17$ than for the control group $<g>_{28C}=0.32\pm0.18$. The treatment group had medium gains while the control group had low gains. The treatment group average gain was 1.0 standard deviations above that of the control group. (23 in Appendix A)

Results for the FCI selected questions

Honors Physics

In this section are the results for the 11 FCI questions that I hoped our graphical treatment would help improve. The honors students had a pretest mean of 4.29 and a posttest mean of 8.79 out of 11.

Class	Test	Pretest mean	Posttest mean	Normalized Gain for pre and posttests
Honors Physics	FCI-selected	4.29	8.79	0.68

There was no control group in honors physics with which to run a comparison therefore I compared the pre and posttests by running the Hake test for normalizing the gain and found it to be 0.68.

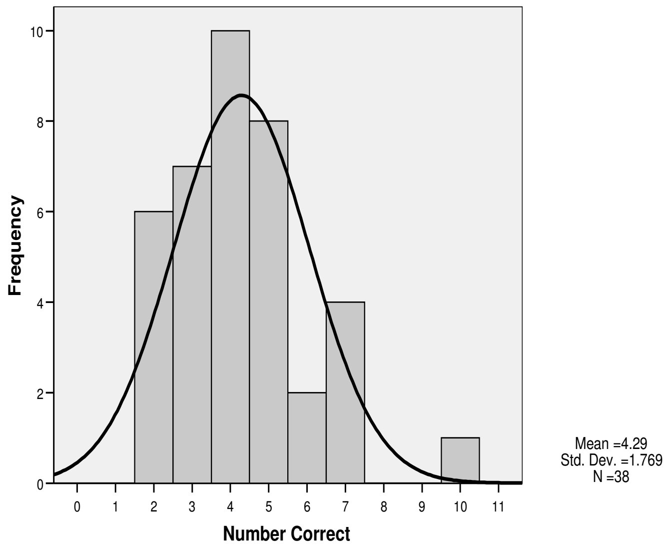


Figure 15. Honors FCI selected questions on pretest

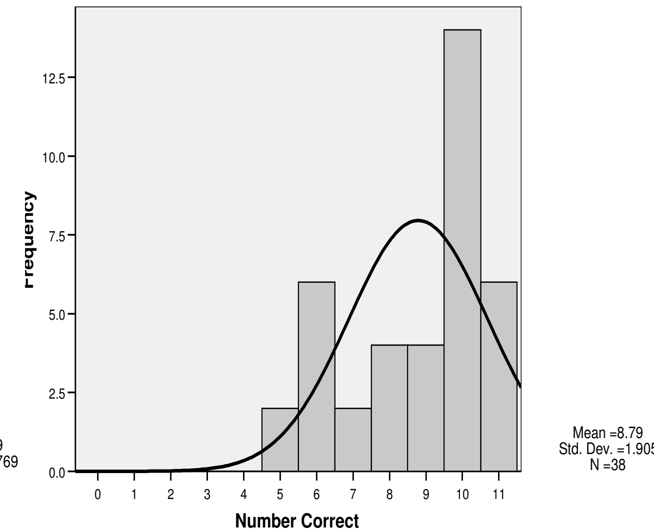


Figure 16. Honors FCI selected questions on posttest

A directional dependent sample t-test was conducted to evaluate the hypothesis that the students in my honors physics class would increase their FCI score on the 11 selected questions. At $\alpha = 0.01$ I reject the null hypothesis that the population mean for the posttest is less than or equal to the population mean of the posttest $t(37)=13.19$, $p<0.001$. I conclude that there is a significant difference in the scores of the students for the pretest (Mean=4.29, SD=1.79) and the scores on the posttest (Mean=8.79, SD=1.91). The results of the t-test indicate that the treatment the honors students received led to higher scores on the posttest. There was a low correlation between the students pretest and posttest scores of 0.35. (27 in Appendix A)

The pre and posttest performance of the honors students on the FCI selected questions was also compared using the Hake test. It was determined that the average gain for the group was $\langle g \rangle_{38}=0.68\pm0.27$. The honors students demonstrated medium gains on the FCI selected questions. (28 in Appendix A)

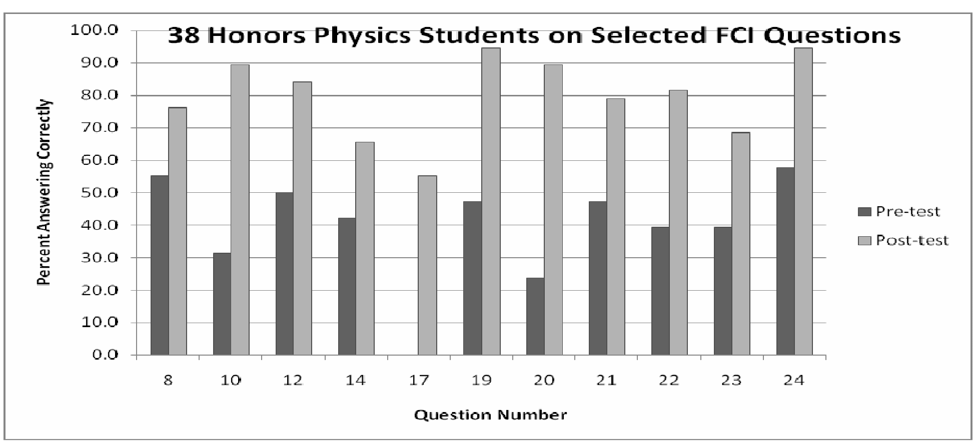


Figure 17. Honors physics treatment group performance on the 11 selected questions in the FCI test

Our group hoped to show that the FCI questions that involved some spatial reasoning would show a marked improvement due to our treatment. I decided to look at the student performance on specific selected questions. Question 17 pertains to an elevator being lifted at a constant speed and determining what force on the elevator is greatest. Interestingly no student answered question 17 correctly on the pretest and on the posttest the question was still answered correctly by the fewest students. The TUG-K test showed the students understood the concept of constant velocity, however the bridge between that and Newton's Laws was shaky for this question. Questions 19 and 20 involved motion of blocks whose positions were shown at regular intervals. These are the only questions for which the correct answer depends only on understanding motion. The students did very well on those two questions.

General Physics

The general physics FCI pretest mean was 7.39 and the posttest mean was 18.68 out of 30 with a normalized gain of 0.50. The control group of general physics students had a pretest mean of 9.40 and a posttest mean of 13.35 with a normalized gain of 0.24.

Class	Test	Pretest mean	Posttest mean	Normalized Gain for pre and posttests
Treatment General Physics	FCI	3.46	7.11	0.46
Control General Physics	FCI	3.72	5.53	0.25

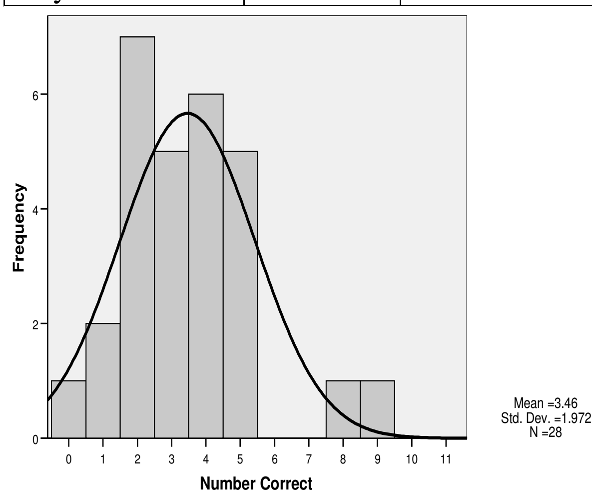


Figure 18. Treatment FCI pretest

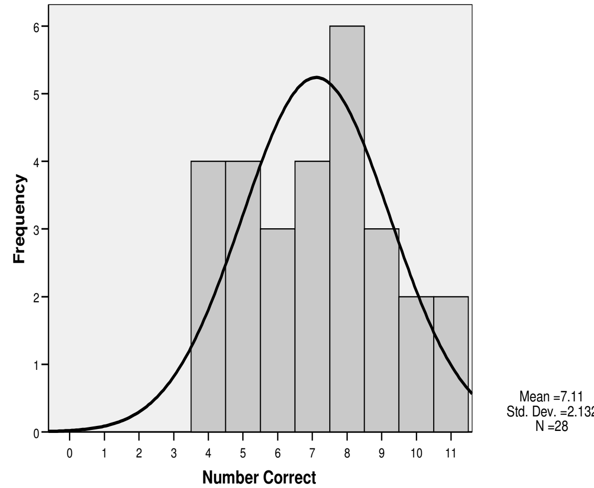


Figure 19. Treatment FCI posttest

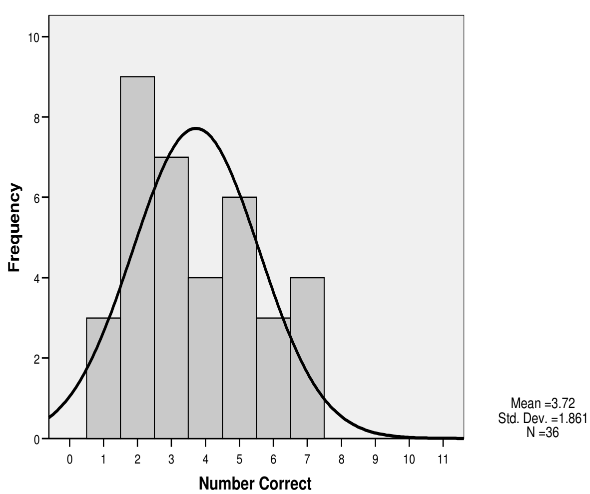


Figure 20. Control FCI selected questions pretest

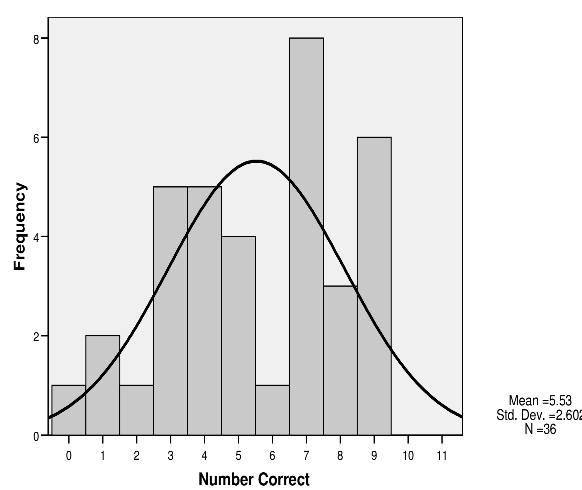


Figure 21. Control FCI selected questions posttest

I was interested in the effect of my graphical treatment on the students' FCI scores. Based on a directional independent samples t-test at $\alpha=0.01$, I reject the null hypothesis that the population of mean treatment students is less than or equal to that of the control group, $t(62) = 2.60$, $p = 0.006$. I therefore conclude that treatment group has significantly higher scores (Mean=7.11, SD=2.13) than the control group (Mean = 5.53, SD = 2.60). (29 in Appendix A)

An analysis was also run on the two populations for the pretest. A non-directional independent samples t-test at $\alpha=0.01$ was run. In this case I fail to reject the null hypothesis that the two populations have the same means, $t(62) = 0.536$, $p=0.594$. The conclusion is that there was no significant difference between the treatment group (Mean = 3.46, SD = 1.97) and the control group (Mean = 3.72, SD = 1.86) at the time of the pretest. (29 in Appendix A)

The performance on selected FCI questions for the treatment group and the control group was compared using the Hake test. Based on the average of the Hake gains of the students I determined there was a substantially larger gain in the treatment group $\langle g \rangle_{28T}=0.45\pm0.40$ than for the control group $\langle g \rangle_{28C}=0.24\pm0.31$. The average gain by the treatment group was 0.52 standard deviations higher than the average gain for the control group. The treatment group had medium gains while the control group had low gains. (30 in Appendix A)

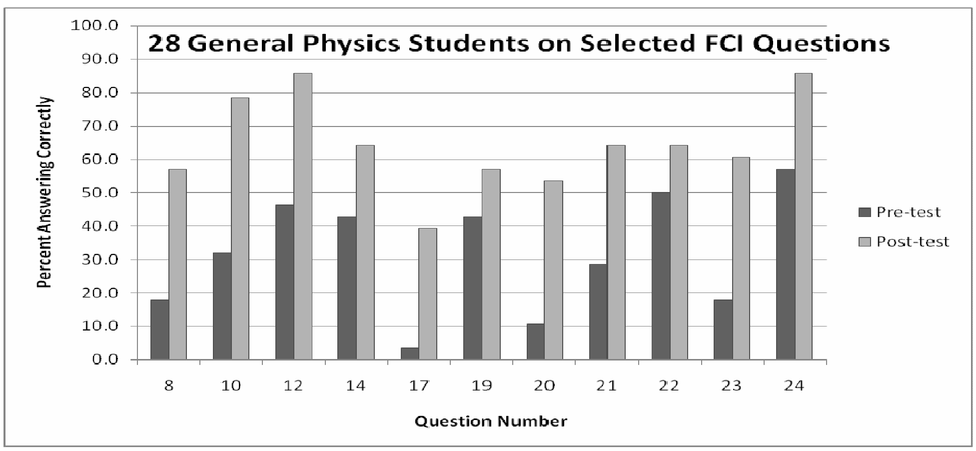


Figure 22. General physics treatment group performance on the 11 selected questions in the FCI test

As with the honors students, the general physics students did not do well on number 17, the elevator question. They did not do nearly as well on questions 19 and 20 involving the motion of the blocks as the honors students. I have noticed for ten years that general physics students do not demonstrate the ability to cast off pre-Newtonian ideas as successfully as the honors students do.

Results of student questionnaires and discussions

I gave the students an exam on unit 3 in which three of the six problems could have been solved either graphically or algebraically. At the beginning of each problem there were axes for each type of motion graph. The three questions are in the appendix. (32 in Appendix A)

After the exam I went through the exams to categorize the method the used to solve the problems. The groupings were; graphical for those that solved all or most using graphs, both for those that solved using both methods, and algebraic for those that used algebra most or all of the time. Later I gave them a questionnaire inquiring as to why they solved the problems using their method of choice. Here are comments from some of the students.

Emily D.: "I did it that way because that's what comes easiest and is the most natural to me."

Emily always used algebra. She never placed anything on the graphs.

Mark I.: "I liked to use Algebra formulas better because they worked out better for me. You just plugged in numbers and found your answer. I believe they are more accurate and they were easier to use for me."

Mark used algebra most of the time but did use graphing to help him on one problem.

Claire L: "I preferred the graphs because I had already had a calculus class and I used the concept of integrating. Plus it was easier for me to just remember that acceleration is the derivative of velocity and that is the derivative of position so then all I need to worry about is basic geometry. Although the equations were also derivatives, it still made more sense to do problems that way."

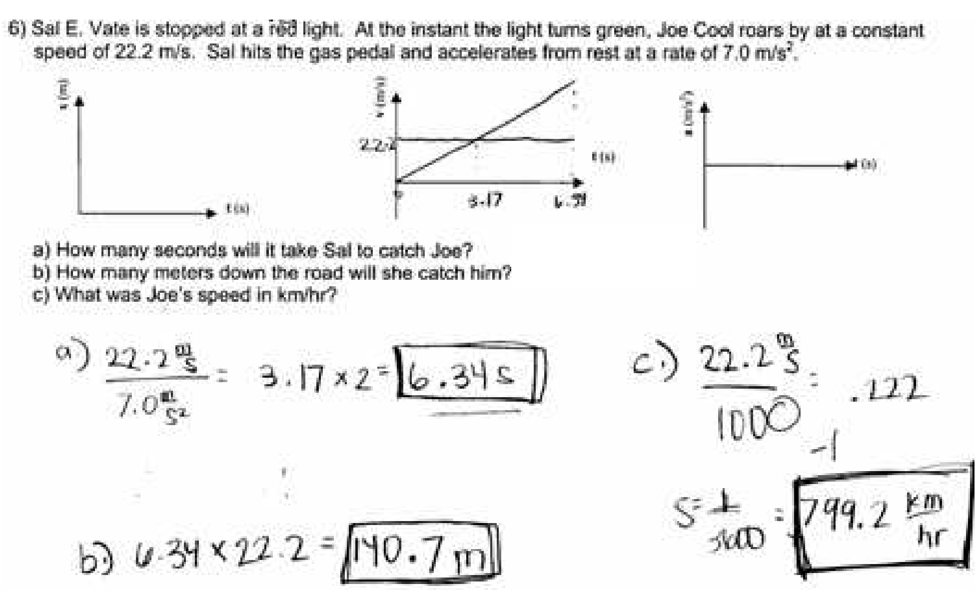
Claire used graphs to solve most of the time. I noticed during the year that she often used both methods to check out her work. A couple of times during the year she wasn't sure about equations and would check them out graphically.

Cassie M.: "I preferred the algebra because the numbers are never wrong and the equations are easier and less room for error than measuring and making graphs."

Cassie did all of her calculations algebraically. However, she always sketched a graph when solving for displacement and divided it into sections. She often sketched graphs in other situations. I asked her why she bothered to sketch the graph when she preferred to use algebra. She said it helped her to think out the problem and what steps she had to include.

Jenna M.: "I prefer to use algebraic formulas over graphs. I have a hard time visualizing graphs. I understood it better if I solved it with algebra."

Interestingly enough, Jenna solved over half of the problems using graphs. I guess to her, doing the problem using the slopes and areas of the graph was algebraic. Below is her setup for a difficult problem.


6) Sal E. Vate is stopped at a red light. At the instant the light turns green, Joe Cool roars by at a constant speed of 22.2 m/s. Sal hits the gas pedal and accelerates from rest at a rate of 7.0 m/s².

a) How many seconds will it take Sal to catch Joe?
b) How many meters down the road will she catch him?
c) What was Joe's speed in km/hr?

a) 22.2 m/s ÷ 7.0 m/s² = 3.17 × 2 = 6.34 s

b) 6.34 × 22.2 = 140.7 m

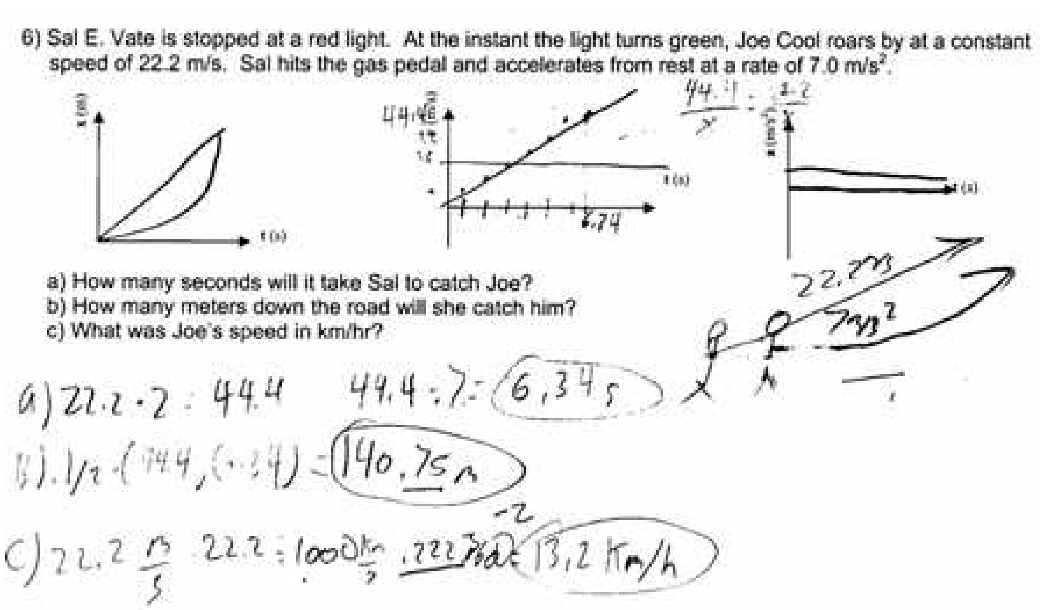
c) 22.2 m/s ÷ 1000 = .222 ; s = 1/3600 : 799.2 km/hr

Shelby M.: "I prefer algebraic formulas because it is easier for me to plug in things I already know and solve for something missing. However sometimes graphically helped me visualize it."

Shelby used algebra most of the time but usually sketched a motion graph and sometimes used a velocity vs. time motion graph to solve for displacement..

Andrew P.: "I used the graphs in this case because the shapes created by the graphs were simple. So the easier the shapes, the easier it was for me to calculate slopes and areas by using geometry. It breaks down into simpler math."

Andrew did almost all of the problems using graphing. Below is his setup for the problem. In my twenty years of teaching I have never had a student solve this problem graphically and this year there were six that did.



Rachel R.: "Algebra just makes more sense to plug in values into equations and when I look at an equation it looks right and I can easily see what I need to do and what values I need or have to solve for."

Rachel used algebra most of the time. However she often sketched in a velocity time graph and in one problem she used it to solve for displacement. When I asked her why she drew graphs when she preferred algebra she said they help her think out the problem.

Tom S.: "I think I use graphs because it is faster for me and the easiest to figure out. Also I make more mistakes when I try to use algebra."

Tom did almost all of the problems using graphs to get his answer.

Cody U.: "I preferred the algebra method because it seemed like the values were more accurate using algebra."

Cody used algebra most of the time but occasionally sketched a graph to help him think out the problem.

Conclusion

My treatment group had higher scores on their FCI tests than the control group. However, I do not attribute that to the graphical treatment instituted for this study. The scores are not significantly higher than previous years on the FCI. The likely cause is that I have been using modeling for ten years and teaching physics for twenty eight years. This state of affairs was not true for the control teachers or for other members of my team. The hypothesis that we could increase the FCI scores on questions involving spatial reasoning was not valid. The gain in those questions lagged slightly behind the overall gain in the FCI for my students. Most of the selected questions tie both motion and forces together making them some of the most difficult.

I believe the treatment I used on my students led to higher scores on the TUG-K exam and an increased overall understanding of kinematics. I think that the TUG-K scores are likely the best overall of my career. I am unable to confirm this belief by looking up past scores because while I have been using the TUG-K for ten years, it has always been a section of a larger exam. The statistic that the honors physics treatment group mean was close to ninety percent correct is quite impressive. A study published by the author of the TUG-K test that discussed the test and the performance of high school and college students on it indicate a mean of around fifty percent is common. (Beichner, R. 1994) By analyzing the TUG-K results one question at a time it is evident that more emphasis needs to be placed on acceleration vs. time graphs and on taking slopes of lines that do not begin at the y-axis.

Even if I had not seen such high TUG-K scores I would still be a proponent of emphasizing graphical methods of solving problems before introducing algebraic methods. My students this past year exhibited a much deeper understanding of the information that can be extracted from a graph than in years past. When my students studied the concepts of impulse and work, the skill of using the area of a graph to solve for those values progressed more smoothly than any other year. I do feel the extra three days spent on emphasizing the skill of solving for values using graphs is worth it and I intend to maintain the practice.

In discussions with the students it was evident that the emphasis on graphing first did not lead to all of the students using graphs to solve their kinematics problems. However many of the students that solved problems algebraically did use graphs to help them set up the problems. Over forty percent of my students used graphs to solve problems or to assist them setting up the problem in the unit 3 problem exam. (31 in Appendix A) This has never been the case in years past. If nothing else, I have emphasized another tool for the students to place in their “toolbox” that may be useful for them.

Angela McClure's Field Trial Report

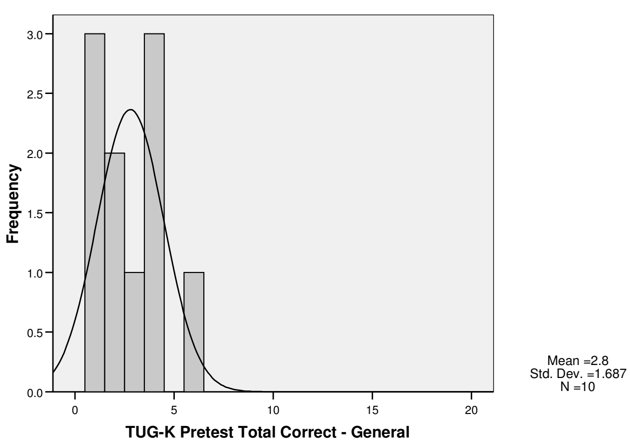
Procedure

My sample consisted of 19 students with varying science and math backgrounds. 10 students were enrolled in a general physics course. 9 students were enrolled in an honors physics course. Students from both courses were concurrently enrolled in an Intermediate Algebra, Precalculus or Calculus math course.

There was very little variation in treatment method from the other investigators. Students were required to solve problems using graphing methods first. After the problem was solved graphically the students were required to solve the same problem algebraically using physics formulas. Students therefore received equal instruction using both graphical and algebraic methods to solve problems. Students were given pre and post tests of the FCI and TUG-K assessments. Over the course of the treatment students were also asked to solve 3 specific problems on thin strip of paper to determine the method in which they chose to solve the given problems. (56 in Appendix A) At the end of the course (May 2008), 6 students were interviewed and video taped while solving strip question 2.

Results

I was interested in the effect of the treatment on the general physics students' kinematics understanding as measured by the TUG-K. Based on a non-directional paired samples t-test at alpha = 0.05, I reject the null hypothesis that population mean of the pretest takers is equal to the population mean of the posttest takers, $t(9) = 7.575$, $p < 0.001$. I conclude that there was a significant difference in kinematics understanding as measured on the Tug-K for general physics individuals before and after the treatment. (Treatment Pretest: Mean = 2.80, SD = 1.687. Treatment Posttest: Mean = 10.90, SD = 3.510). I was 95 % confident that the interval -10.519 and -5.681 contains the true population mean difference. The correlation was 0.315. (52 in Appendix A)



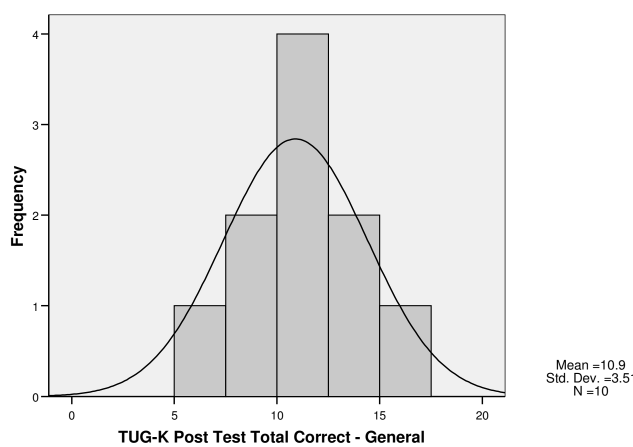


Figure 1. Comparison of histograms for McClure's General Physics Sample Tug-K pretests (left) and posttests (right).

At the beginning, these students had some of the lowest scores in the study as can be seen from the left centered graph of pretest scores. The second graph shows a shift toward the center or right of the graph. From these results, it can be concluded that the general physics students who received the treatment showed significant gains in their understanding of kinematics concepts.

I was also interested in the effect of the treatment on the honors physics students' kinematics understanding as measured by the Tug-K. Based on a non-directional paired samples t-test at alpha = 0.05, I reject the null hypothesis that the population mean of the pretest takers is equal to the population mean of the posttest takers, t(8) = 22.235, p < 0.001. I conclude that there was a significant difference in kinematics understanding as measured on the Tug-K for honors physics individuals before the treatment and after the treatment. (Treatment Pretest: Mean = 5.78, SD = 2.224. Treatment Posttest: Mean = 15.89, SD = 1.900). I am 95 % confident that the interval -11.160 to -9.062 contains the true population mean difference. The correlation was 0.792. (46 in Appendix A)

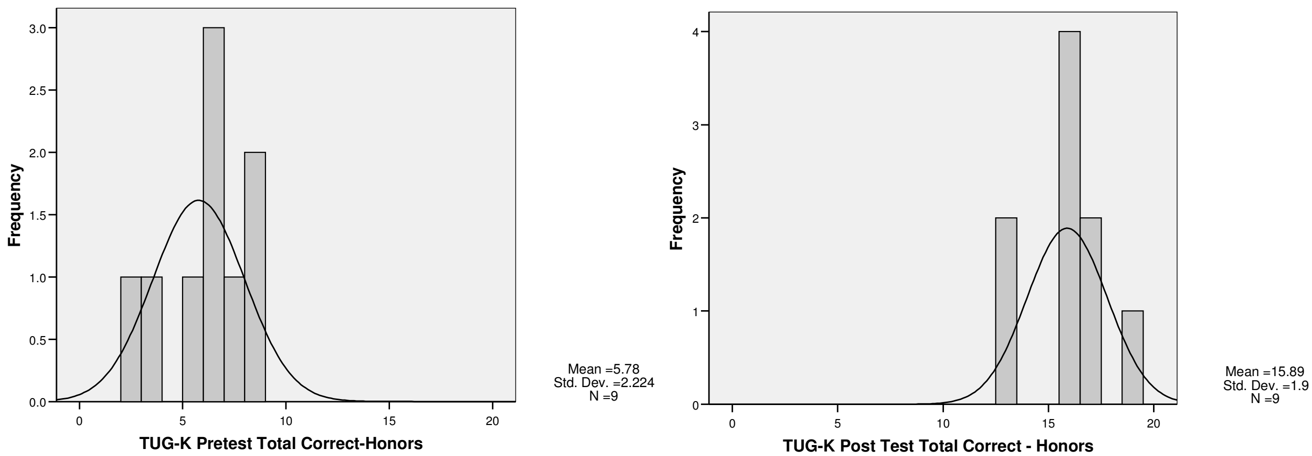
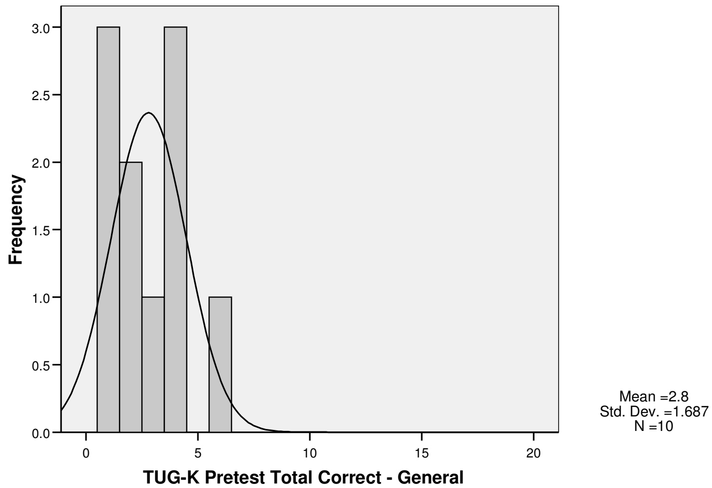


Figure 2. Comparison of histograms for McClure's Honors Physics Sample Tug-K pretests (left) and posttests (right).

The above graphs show TUG-K scores that shifted considerably from the left section of the graph to the right. From these graphs and t-test results, I conclude that the honor students who received the treatment also showed significant gains in their understanding of kinematics concepts.

My intention was to compare the gains of the general treatment group to that of the control group. To further this goal, I was interested in determining whether the treatment samples and the control samples from the Tug-k pretest were from the same population. Based on a non-directional independent samples t-test at alpha = 0.05, I reject the null hypothesis that population mean of my general physics treatment group is equal to the population mean of the control group, t(42.894) = 7.024, p < 0.001. I conclude that there was a significant difference between the treatment sample and the control sample on the Tug-K pretest. (Treatment Pretest: Mean = 2.80, SD = 1.687. Control Pretest: Mean = 9.40, SD = 4.893) I am 95 % confident that the interval 4.705 to 8.495 contains the true population mean difference. (53 in Appendix A)



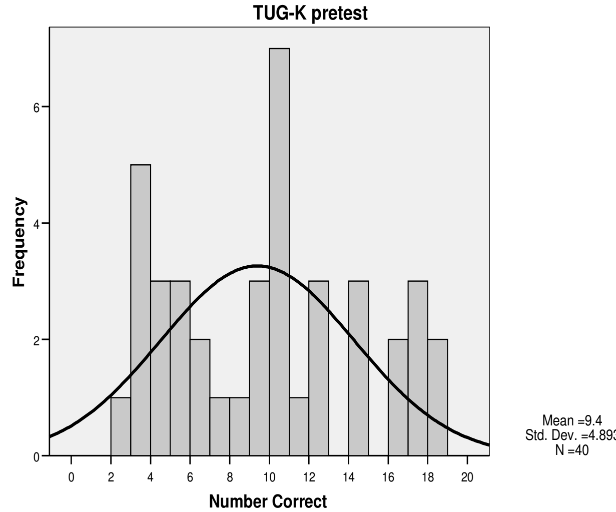


Figure 3. McClure's Treatment TUG-K pretest histogram on the left. Control TUG-K pretest histogram on the right.

From these results, I conclude that the general physics treatment group and control group who took the Tug-K pretest did not come from the same population, and a traditional statistical comparison between these samples' scores on the Tug-K using further t-tests on this test is inappropriate.

Hake Gain Test

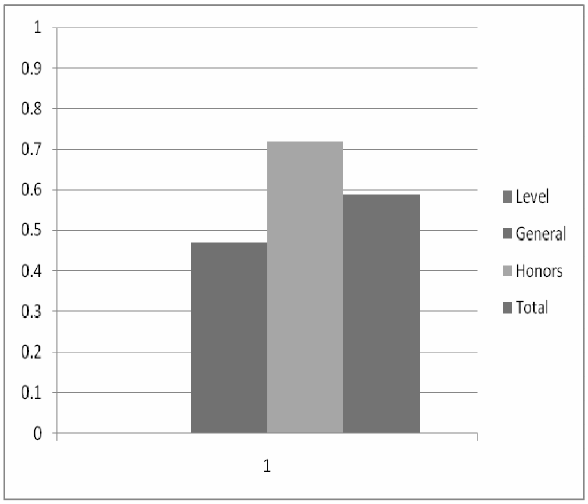
Comparing the treatment and control group required a more appropriate method of comparison than a traditional t-test. I then determined to use the Hake Gain test. (33-34 in Appendix A) The Tug-K Hake Gain results for the treatment group are as follows:

McClure's Treatment Group	General Physics (N=10)	Honors Physics(N=9)	Total Sample (N=19)
Mean Hake Gain Score	.471	.718	.588
Standard Deviation	.199	.103	.201

Figure 4: Mean Hake Gain Scores for McClure's Treatment Group

CONTROL	General Physics (N=10)
Mean Hake Gain Score	.245
Standard Deviation	.517

Figure 5: Mean Hake Gain Scores for the Control Group



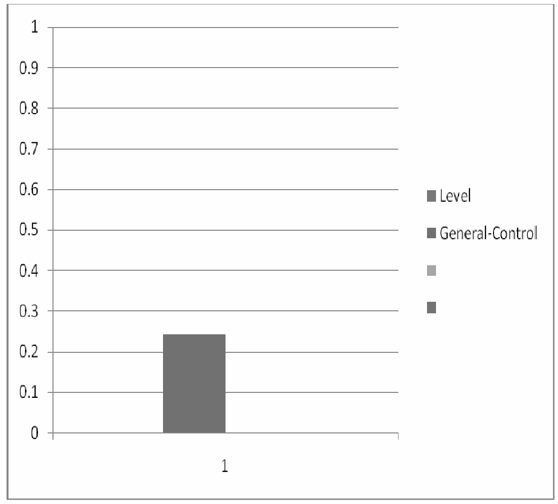


Figure 6: Mean Hake Gain Scores for McClure's Treatment Group(s) on the left. Mean Hake Gain scores for the Control Group on the right.

The results show that my general treatment was 1.14 standard deviations above that of the control group. My honors sample was 4.60 standard deviations above that of the control group. My combined sample was 1.71 standard deviations above that of the control group. In conclusion the Hake Gains were significantly greater in the treatment sample than in the control sample. Considering both that the treatment group began at a substantially lower level than that of that of the control group, and yet had much larger Hake gains, it can be concluded that treatment had an impact on the treatment sample.

Strip Questions:

I was also interested in the effect of the treatment on the method in which the students used to solve problems. A frequency distribution was used to determine the percentage of students that used graphing, algebra or a combination method to solve each problem. The results are as follows:

STRIP QUESTION #1

Total Students in Treatment Group (N=18)

Method	Frequency	Percent
Graphical Method	12	66.7
Algebraic Method	3	16.7
Combination of Graphical and Algebraic Method	3	16.7

Figure 7: Frequencies of solving methods for Strip Question 1 for the McClure Treatment Group.

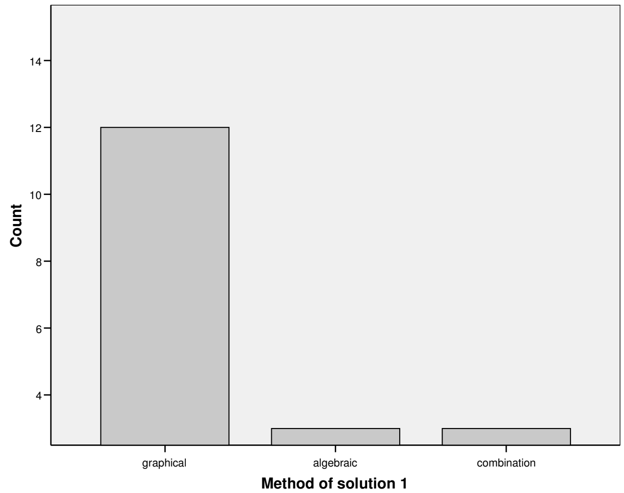


Figure 8: Histogram of solving methods for Strip Question 1 for the McClure Treatment Group.

STRIP QUESTION #2

Total Students in Treatment Group (N=19)

Method	Frequency	Percent
Graphical Method	9	47.4
Algebraic Method	2	10.5
Combination of Graphical and Algebraic Method	8	42.1

Figure 9: Frequencies of solving methods for Strip Question 2 for the McClure Treatment Group.

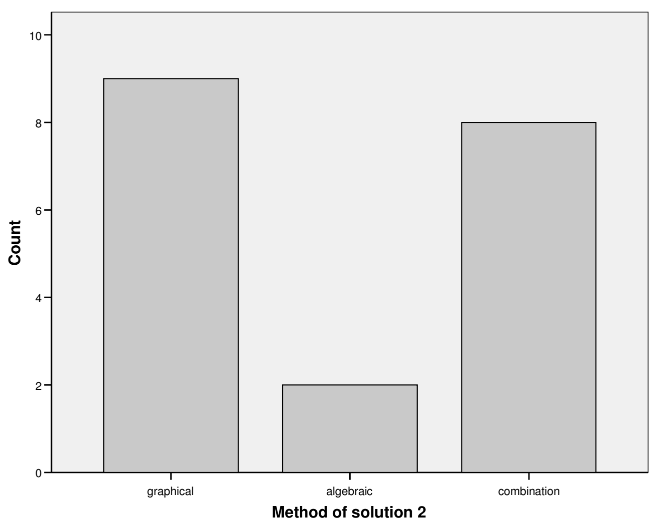


Figure 10: Histogram of solving methods for Strip Question 2 for the McClure Treatment Group.

STRIP QUESTION #3
Total Students in Treatment Group (N=17)

Method	Frequency	Percent
Graphical Method	15	88.2
Algebraic Method	1	5.9
Combination of Graphical and Algebraic Method	1	5.9

Figure 11: Frequencies of solving methods for Strip Question 3 for the McClure Treatment Group.

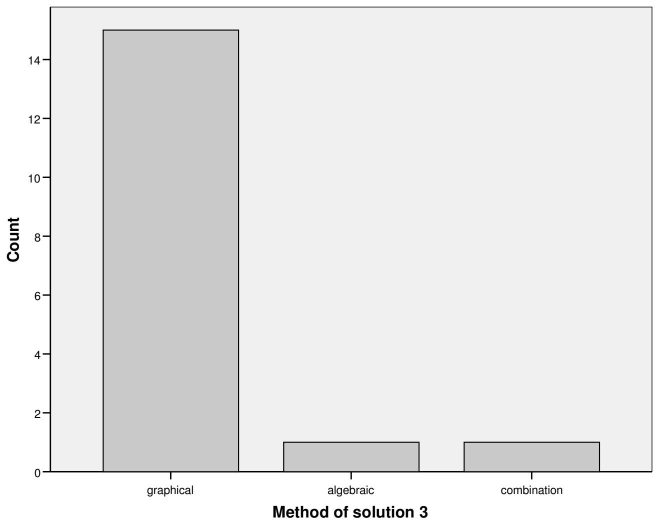


Figure 12: Histogram of solving methods for Strip Question 3 for the McClure Treatment Group.

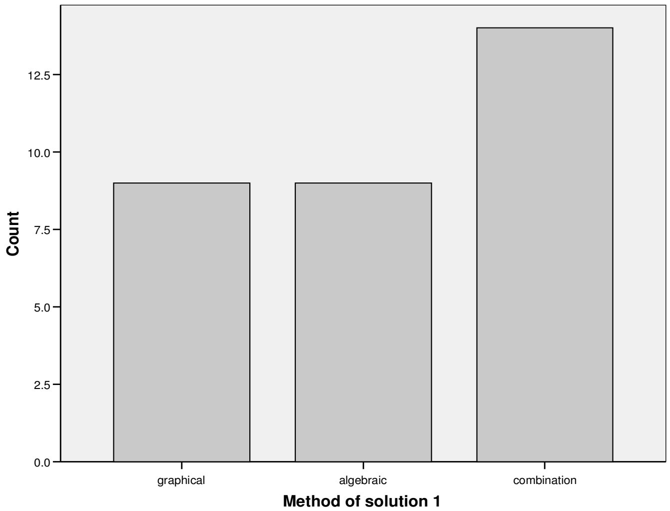
The 3 strip questions were given over the time of the treatment. The percent of students who used algebra on the first strip question was 16.7% . The percentage of students who used strictly algebra to solve then decreased to 10.5% for question 2, and then additionally decreased to 5.9% by the final strip question. If the number of students using algebra to solve decreased over time, then the amount of students using some form of graphing method to solve then increased over time. Thus it appears that students in the treatment group chose some method of graphing over a strictly algebraic method. These results seem to follow the trend established by the entire treatment sample. (36-43 in Appendix A)

The Control Group reported results only for Question #1 and Question #2. The results are as follows:

STRIP QUESTION #1
Total Students in Treatment Group (N=32)

Method	Frequency	Percent
Graphical Method	9	28.1
Algebraic Method	9	28.1
Combination of Graphical and Algebraic Method	14	43.8

Figure 13: Frequencies of solving methods for Strip Question 1 for the Control Group.



McClure Figure 14: Histogram of solving methods for Strip Question 1 for the Control Group.

STRIP QUESTION #2
Total Students in Treatment Group (N=41)

Method	Frequency	Percent
Graphical Method	4	9.8
Algebraic Method	37	90.2
Combination of Graphical and Algebraic Method	0	0

Figure 15: Frequencies of solving methods for Strip Question 2 for the Control Group.

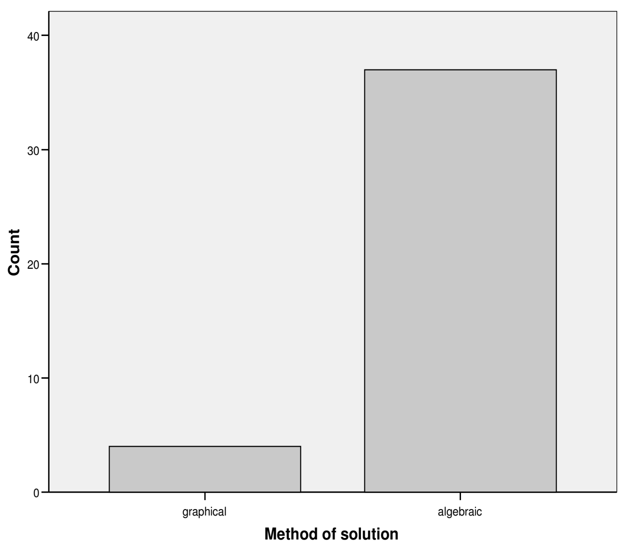


Figure 16: Histogram of solving methods for Strip Question 2 for the Control Group.

Very few of the students in the treatment group used graphing to solve the first problem, but by the second problem, most of the students used graphing to solve. In the control group I saw the exact opposite trend. Most of the control group students solved the first problem by graphing, but when the problems got harder, they reverted to solving the problem with algebraic methods. These observations follow the trend established by the entire treatment sample. (36-43 in Appendix A)

Investigation of Strip Question 2

The following figures are examples of student work for strip question 2. Both students answered the question correctly. One student solved using a graphing method, while the second student solved the same question algebraically. Both students solved the question correctly; however, the graphically method demonstrates a much simpler mathematical process.

A scan of four whiteboards was deleted to make the document a smaller file.

Figure 17. Student work on the left represents a correctly answered question solved graphically. (Treatment Group) Student work on the right represents the same correctly answered question solved algebraically. (Control Group)

The following work illustrates two students' work for strip question 2. Both students answered the question incorrectly. One student solved using a graphing method, while the second student solved the same question algebraically. Both students had difficulty finding the correct time interval; however, the student who solved the question graphically shows a deeper understanding of the difference between constant velocity and acceleration. The student on the left did make a connection that the time interval for acceleration portion must come from subtracting the initial 2 seconds from the final time, but continues to use the incorrect time interval throughout the problem. The student then continues to make small mathematical errors, but shows a basic understanding of the physics concepts. The student who solved algebraically did not make the connection that the displacement formula they had used only applies to an accelerating object. Thus the student solved the problem mathematically correct, they however had difficulty applying it with the larger physics concept.

A scan of whiteboards was deleted to make the document a smaller file.

Figure 18. Student work on the left represents a incorrectly answered question solved graphically. (Treatment Group) Student work on the right represents the same incorrectly answered question solved algebraically. (Control group)

Student Work

Students were asked to find the time a ball reaches its highest point if tossed up at 75m/s. The students were also asked to find the height of the ball when it reaches its highest point. The following work was taken from 3 of 5 groups. All 5 groups solved the problem using the following method. Students first started with motion maps, x-t, v-t, and a-t graphs. Most students realized that the velocity at the ball's highest point was 0 m/s. They struggled with the unknown time of the ball at the highest point. All 5 groups decided to make a list of velocities each second. They continued the process until the ball reached -5m/s at 8seconds. Students' list didn't contain a velocity of 0 m/s. Students then reasoned that the ball reached its highest point between 7 and 8 seconds. Students determined that because the ball is accelerating at a constant rate, the ball reaches its maximum point at 7.5 seconds. Students then used the v-t graph and the area under the curve to find the displacement of the ball.

Students used strictly logically reasoning and what they understood conceptually of acceleration. Students did not use a single physics formula. If this problem was to be solved algebraically it would have required several formulas and substitutions. This was a cooperative learning activity and students worked in teams to solve the problem. All 5 groups solved the question with the same method and solved it correctly.

A scan of \whiteboards was deleted to make the document a smaller file.

Figure 19: Example of students' whiteboards from both Investigator 4's General and Honors Physics.

Interviews

In order to investigate the strip questions more thoroughly question #2 was given to students at the end of the school year. (6-7 months after the material in question #2 was presented.) Select students were asked to attempt the question again and were videotaped and interviewed. The purpose of this procedure was to gain insight to how students reasoned through the questions and the methods in which they used to solve the problem. The interview was done later in the school year to gain insight into what helps students to retain information long term. (54 in Appendix A)

The three students interviewed from the general physics course at first complained about not remembering "any formulas." Students were then asked if there was anything they could do to stimulate their memory and all three chose to graph. Although they were led a bit, they all continued to use graphing throughout the entire interview. All three students solved for displacement by using the area under the v-t graph. Two of the three students solved the entire displacement portion correctly. The student who did not solve the problem correctly solved for the displacement while the car was slowing, rather than both when the car was slowing and traveling at a constant velocity. All three students had no issue separating the constant velocity portion of the graph from the acceleration portion. In past years my students had issues remembering when to use the constant velocity formulas versus the acceleration formulas. I was pleased that the treatment seemed to help the students distinguish between the two concepts.

The three students solved for the magnitude of the acceleration correctly. One student needed coaxing whether to multiply or divide to calculate acceleration. After some questioning she used her graphs and the area under the a-t graph to work backwards to solve for acceleration. This was particularly interesting because this method was taught briefly over 6 months earlier. The student intuitively used a method that was not emphasized in the course. The other two students used the slope concept to solve for acceleration; however, they could not make the connection to the word slope.

All three students found the correct time interval easily and drew the v-t graph with a decreasing line. In past years if my students solved this question algebraically they would mistakenly use the entire time interval rather than only the acceleration portion. Although the students drew the graphs correctly all three students at first did not put negative signs on their answers. It was clear that conceptually the students understood the concepts but had problems executing the final answer perfectly. As the interview progressed two students caught their mistake and marked their acceleration with the correct sign.

The three students interviewed from the honors physics course had no issue with this question. All three students drew the graphs immediately and solved all parts of the problem correctly in just a few minutes. One student made a few mistakes because he hastily solved the problem. After he was encouraged to slow down and look carefully at the problem he solved the problem beautifully.

Students were asked, “Why didn’t you use physics formulas to solve?”

Student responses: “I couldn’t remember the formulas, but I could remember the pictures.”
“Because that is what you taught us to do.”
“It is easier and faster to graph.”

Students who were concurrently enrolled in calculus were asked, “How has graphing helped you?”

Student responses: “There were multiple questions on the AP Calculus Exam that I could solve graphically.”
“It helped me to understand derivatives and integrals.”
(55 in Appendix A)

Conclusion

Evidence shows if nothing else that the students who received the treatment felt as or more comfortable with interpreting and using graphs than otherwise without the treatment. After viewing the whiteboarding sessions, it is evident that the treatment gives students a visual reference to communicate from. It helped students to follow their fellow classmate’s logic visually rather than only verbally and mathematically. The treatment made a difference in the students’ problem solving skills. It also convinced many students that visual representations are important for learning and necessary to communicate with others.

In addition, calculus level students gained confidence and made connections with their calculus material. The treatment allowed for a smooth transition to other sections of the course where a connection to graphing and area under a curve could be made.

I am committed to continue and expand on the treatment in my future courses. Students in this particular group began at a significant lower level than that of the control group and yet they made greater gains on the TUG-K. Many lower level students gained confidence and enjoyment. Solving graphically has allowed students with lower math ability to make gains and participate with confidence among their fellow classmates.

Implications for Further Research

Though this research was successful in finding a difference in kinematics understanding as measured by the TUG-K in the treatment group versus the control group, it is based on relatively small sample sizes and a control group that did not match the population of the treatment group. Further study of the effect of solving problems graphically prior to introducing algebraic solutions for kinematics problems is necessary to determine the overall size and depth of student improvement with this method. Though we believe that the effect that we found is real, more work should be done to determine how large an effect there would be with a larger population of students.

Acknowledgements

When we embarked on the task to complete this research, we did not really know what we were doing or what we were getting ourselves into. This entire endeavor was a learning process, and there were many people who assisted us along the way. We would like to thank the following individuals for their support and advice. We thank **Dr. Colleen Megowan** for her guidance and advice during the research and writing of this paper. We thank **Jane Jackson** for all of the financial support that she provides to the MNS program and to the teachers individually through the pursuit of grants and all of the work she does to keep the MNS program going. We thank **Dr. Robert Culbertson, Dr. Carl Covatto, and Dr. Michelle Zandieh** for working with the MNS degree program to improve physics education for ourselves and the students that we teach. We thank **Dwain Desbien** for "planting some seeds". We thank **Meegan Addy, Nathan Justis, Bill Lewis, and David Meyer** for allowing us to gain from the wisdom of their experience. We would like to thank **Karen Burrington** for helping us solve the innumerable problems that popped up along the way. A special thanks to **Greg Stafford and Erin Hartse** for giving up class time to conduct numerous assessments. Finally, we thank **our families** for the sacrifices that they made during the long hours and unexpected delays that came up as we worked days, nights and weekends to complete our research.

Works Cited

Baartmans, B. G., Sorby, S.A. (1996) Introduction to 3-D spatial visualization (and accompanying teacher's resource manual, Prentice Hall

Beichner, R. (1994). Testing student interpretation of kinematics graphs. American Journal of Physics, 62, 750-762

Brown, J., Collins, A., Duguid, P. (1989, Jan.-Feb.) Situated Cognition and the Culture of Learning. Educational Researcher, 18(1), 32-42

Hake, R., (1998, January). 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

Hake, R. R., (2002, August) Relationship of individual student normalized learning gains in mechanics with gender, high-school physics, and pretest scores on mathematics and spatial visualization. Submitted to "The Physics Education Research Conference"; Boise Idaho

Hestenes, D., (1987, May). Toward a modeling theory of physics instruction. American Journal of Physics, 55(5), 440-454

Hestenes, D., Wells, M., Swackhamer, G. (1992, March). Force Concept Inventory. The Physics Teacher, 30, 141-158

Jackendoff, R. (1996). The architecture of the linguistic-spatial interface.
In P. Bloom, M. A. Peterson, L. Nadel M. F. Garrett (Eds), Language and space (1-30). Cambridge, MA: MIT Press.

McDermott, L. (1993, April) Guest Comment: How we teach and how students learn - A mismatch? American Journal of Physics, 61(4), 295-298

McDermott, L. Rosenquist, M. (1987, June). Student difficulties in connecting graphs and physics: Examples from kinematics. American Journal of Physics, 55(6), 503-513

Megowan, M. (2007) Framing discourse for optimal learning in science and mathematics. (Doctoral dissertation, Arizona State University, 2007).

Below are the two strip questions used in the action research called

The Effects of Developing Kinematics Concepts Graphically Prior to Introducing Algebraic Problem Solving Techniques

By James Archambault, Theresa Burch, Michael Crofton, Angela McClure
Arizona State University, July 2008.
Available at http://modeling.asu.edu/Projects-Resources.html

Test questions:

Constant velocity:

Ann T. Matter travels in the negative directions at 20 m/s for 10 s, after which she instantly changes to 30 m/s in the positive direction and moves for 5 s, after which she instantly changes to 10 m/s in the negative direction and moves for 10 s, and finishes by moving in the negative direction at 5 m/s for 5 s.

What was Ann's displacement for the trip?

Constant acceleration:

Ella Vader is driving her flashy Porsche down the highway at 30 m/s. She sees a deer ahead and 2.0 s after she spots the deer she hits the brakes. It takes her 7.5 more seconds to stop.

How far did Ella travel before stopping?

What was Ella's acceleration while she slowed down?