Learning in the Laboratory: Does Group Work Work?

John Belanger

This paper was completed and submitted in partial fulfillment of the Master Teacher Program, a 2-year faculty professional development program conducted by the Center for Faculty Excellence, United States Military Academy, West Point, NY, 2015.

Abstract

A cadet complaint about group work in my CH459 Chemical Engineering Laboratory course led me to think about the challenges and virtues of group work and seek out the guidance of experts in the form of a literature review. This exploration has made me realize that many knowledgeable researchers still debate whether group or individual effort results in the best outcomes, but that group work definitely offers advantages and is almost certainly a more effective method for structuring my lab course. While group work will remain a fixture in CH459, educator recommendations offer ways to improve the incentives for teammates to produce good group work by allowing teams to self-assemble after the first lab iteration. Providing meaningful peer-evaluation feedback that actually counts in some measure toward individual grades would also improve my course. Reading studies in academic settings caused me to realize that my course is a good environment for such research of my own to help determine whether or not group work actually does produce something greater than the sum of individual efforts. It is a question that remains unresolved.

Introduction

“I hate group work,” one of the top performing students in the chemical engineering major told me at the end of the semester last year. He went on to say that he always ends up doing the lion’s share of the work while everyone, including non-contributors, gets the same amount of credit and he does not learn from the other members. I tried to spin this and put things in a positive light, that perhaps the other students are learning a lot from him and that he is learning a life lesson about how to motivate others to get work done, a key skill for a cadet at West Point who will soon become an Army leader. Having said that, a part of me wondered if he had a valid point about group work, in general, being carried out at universities across the country and in particular, a point about how my course should be run, with individual efforts receiving more weight in grading.

In the business world, presumably, deadbeat employees are fired. In academic group work, there is the potential problem of the free-loader/free-rider too, but usually the team does not have the same options available that their corporate counterparts do. “Firing” another member of a team who does not contribute would most certainly have social implications (likely negative for those doing the firing more so than the person being fired), but it usually would not
help any member of the team receive a better grade. In the business world, there would be a larger portion of the profits to divide among the remaining employees. In academia, it seems there often are few repercussions for deadbeat team members.

This one conversation about group work pushed me to look into what scientists, psychologists, economists, game-theorists, and educators – in short, experts from a variety of disciplines – have discovered about the effectiveness of learning done in groups and methods of fairly assessing group work. I wanted to know if there are ways to make group performance better than individual work, if groups increase student motivation to learn, or if group work really works at all. I uncovered a lot of debate and no conclusive smoking gun evidence that either group or individual work is better for producing a better final product, but did reach the conclusion that with the setting of my course in a laboratory, team projects do produce a better, safer learning environment. A better incentive and evaluation system could make group work fairer and reduce the amount of “hate” expressed by students who carry more of the burden than the average member does in group projects.

Background

I teach CH459, Chemical Engineering Laboratory, at the United States Military Academy in West Point, New York. This is a course taken by undergraduate chemical engineering students in their senior year (who are called “firsties” in West Point parlance), and it serves as a culminating experience, where concepts developed over the previous three years in other courses are brought to bear over the course of the semester with six different experiments in the lab using different pieces of equipment, such as a cooling tower, heat exchanger, and other fundamental industrial units, so called “unit operations” in chemical engineering.

As course director, I have control of the sections and course material (and really everything that occurs in the course within the guidelines of the Dean, Department Head, and Chemical Engineering Program Director). As the instructor for the two sections of eleven and twelve students each last year, I assigned cadet teams with three or four people in each group. I based the team assignments on the grades these cadets had received in the chemical engineering major prior to beginning their firstie year and starting my fall laboratory course. I ranked the cadets by their grade point averages (GPAs) in the major and then numbered one to three repeating down the list of names to get three groups in each section. This put high, medium, and low GPA students together. My thought in doing this was to make sure that each cadet team had roughly equivalent skill levels and there were no “hero” or “zero” teams that would complete the projects much more slowly or quickly than other groups.

Grades in the course were based on several large individual assessments (a mid-term exam, final exam, and online safety training certification) and also six group projects. The exams were worth a significant amount of points toward the final grade, but group work did account for the majority of the points in the course (53% of the grade, to be exact). Each group
turned in their lab homework and lab projects to me, and everyone in the group received the same grade.

At the end of each experiment, cadets completed a multi-disciplinary team assessment of the group leader, and the group leader completed an assessment on one group member of their own choosing. These ended up all being focused on a team member who stood out for especially good or bad performance, almost always for good performance. Of the eighteen peer assessments that leaders completed on members of their teams over the semester, only two leaders were ever critical of a team member. The feedback for group leaders from their peers was more variable.

The peer evaluations were not graded in any way, but instead I collected the forms, which were essentially rubrics where cadets gave their peers a 1 (worst) to 5 (best) rating in technical competence, communication, organization, and teamwork. There is also a section for writing in comments. Upon receiving all the evaluations, I met with each team leader and standout cadet for one-on-one counseling to give them the feedback that their peers had generated. Sometimes these discussions were useful, but by the end of the semester I got the sense that cadets were just checking the block to fill out the forms and not really thinking about their evaluations.

A random selection of the multi-disciplinary team assessments would most often return a form that has a 5 rating for all of the assessment areas and a simple one-line comment, such as “Good at delegating tasks.”

The actual group project grade was based solely on the content of the overall report or poster, depending on the rotation. Cadets completed four long-form lab reports and two posters similar to the research posters that they present at American Institute of Chemical Engineering poster competitions, a format familiar to most of the students.

Troubles and Questioning

It troubled me that one of the top cadets in the program would say something like “I hate group work.” I also felt that the multi-disciplinary team assessment tool was not being taken seriously or given enough attention to add much to my counseling sessions with cadets. My observations during experiments in the lab provided far more detail than what cadets were willing to document about their peers.

The most infuriating problem occurred during the course when I caught different cadets off-task, playing on iPads or phones in the lab, while their teammates were diligently working. I would have expected peer-pressure alone would have been enough to avoid this problem and no external chastisement would have been necessary. I assumed incorrectly.

These issues were on my mind while I was reading an interesting book that challenges much of the present-day, “common knowledge” about the effectiveness of teamwork. The book
is *Quiet: The Power of Introverts in a World That Can't Stop Talking* by Susan Cain. [1] Cain posits that modern society, the American business culture particularly, has placed too much value in teamwork and extroversion, and that introverts in quiet spaces often achieve the best results working alone. [1]

So these thoughts made me start to wonder and ask questions about the setup of my course and about working in teams altogether. Does group work actually work better than individuals working separately? Do students learn more effectively if they work together? Does peer pressure motivate students to perform more than laziness motivates cadets to slack and let their hard-working peers do all the work? Is there a better way?

To answer these questions, I started researching and came across some interesting, but conflicting ideas.

**The Assembly Bonus Effect Debate**

I ran across two groups of psychologists who had vastly different ideas about the effectiveness of teamwork. The two schools of thought present arguments reminiscent of the two general viewpoints presented in Cain’s book [1], but with more statistical analysis. Up front, I can say that both studies are compelling, but the statistical analysis involved might serve well as examples in an updated edition of Darrell Huff’s *How to Lie with Statistics*. [2]

**Teams Do Work**

On one side of the debate is a team of researchers and management educators including one Dr. Larry K. Michaelsen. Michaelsen is a firm believer that team-based learning works, and works much better than individual effort. [3]

Michaelsen refutes the idea of others that the upper limit of a team’s performance is based upon the limits of the most competent group member, finding that in 222 groups, the team outperformed the best scoring individual member 97% of the time. [3] The groups consisted of three to eight members, and the aggregate performance of these groups was better than any individual or the sum of all individual efforts. This synergistic increase has been dubbed the “Assembly Bonus Effect.” [3]

While Michaelsen does acknowledge that other researchers have not been able to provide empirical evidence of the Assembly Bonus Effect to corroborate his team’s work, he argues that “The lack of empirical support for the superiority of group decisions may be due, in large part, to the artificial nature of the groups, tasks, or settings in which the research has been conducted.” [3]

To me, this artificiality presents a problem because in higher education, often groups and problems are completely arbitrary, decided upon at the whim of the course instructor. I believe that Michaelsen really has the business world in mind. The problems there are more concrete
and measureable in hard dollars and cents, and intangibles causing subjective judgments about what is important and what should be evaluated are presumably less prevalent.

Assuming that classrooms can effectively emulate the “real-world” experience may make the group work seem less artificial. Michaelsen removed artificiality by presenting problems to groups that were real problems that they might encounter in the workplace. The groups had to make decisions about how to handle the situation, for which there were answers that were more correct than others. The individual members then had to make decisions. The evaluators then compared how often groups versus individuals came to the right decision in the expert’s mind and groups were found to do better than any individual alone.

The answers were scored and the average group score was 89.9 points out of 100, whereas the average of the best individual scores in each group was 82.6, and the average of all individual scores was 74.2. From these scores it appears obvious that something about working in a group causes the sum of the individual efforts in a team to add up to more than the sum of the individual parts, exceeding the work of the best performing individual. Group work does work!

**Team Work Does Not Add Up**

Other researchers delved into the Michaelsen study and published papers addressing that study specifically and discrediting the statistical methods used, which were critical in supporting the argument and final conclusion.

Three years after the Michaelsen study, Tindale and Larson looked into Michaelsen’s work. What they found was that although they could replicate the overall results in a real-world academic setting using quizzes and also replicate Michaelsen’s results by conducting computer simulations using statistical and game-theory methods, the results did not prove that the Assembly Bonus Effect existed.

That sounds confusing, but the real-world example of the student quizzes makes the issue clearer. The students first completed a multiple-choice quiz on their own. They turned this quiz in and then assembled as a group to do the same quiz again. The researchers found that for each question, the group as a whole never got the answer correct unless at least one person got the answer correct as an individual beforehand. If no one got the answer correct before, coming together as a group did not help bring out the correct answer. Essentially, for each question, the group performance was only as good as the best individual performer. The groups performed better on the aggregate set of questions in a quiz because that best individual performer could change for each question.

Another team of educators agreed with this conclusion that teams simply aggregate existing knowledge and do not outperform any one individual within the team. Despite this,
Besedes et al. still recommended group work as a way to spread aggregate knowledge across the team and increase performance overall as a group. [5]

From here, the question becomes whether or not a group chemical engineering lab project is more like an aggregate collection of many questions that different students could each contribute knowledge toward or if a lab project is more like a single task, in which case Tindale would predict that performance would be limited to the best individual work.

**Testing the Assembly Bonus Effect**

The differing ideas about whether or not the Assembly Bonus Effect has been demonstrated leads me to some ideas for future research of my own that could help me determine whether or not the Assembly Bonus Effect is being demonstrated in group work in the lab.

Currently, teams turn in a pre-lab homework as a group and then the final project as group. There is no individual turn-in and no individual grade. In my judgment, the components of the pre-lab homework (three or four questions) are sufficiently different to be considered separate tasks that are aggregated into one larger task, similar to several questions on a quiz. I believe this is also true for the final lab project assignment. There are various components, which even though all relate to the same piece of equipment and same experiment, are vastly different types of problems.

Constraints in the classroom, mainly time and access to lab equipment, would make it highly problematic to have cadets conduct lab projects as individuals and then repeat the process as a group and turn in a group project. The cadets would surely riot. I, as the instructor, probably would riot too, with all the lab reports to grade.

What would be more feasible and also serve as an assessment of group learning would be to have cadets complete the pre-lab homework as individuals for as a graded event and then turn in a group version at some later date. The mid-term and final exams could serve as data points to demonstrate if individual performance improves with this technique. Because my course is split into two sections, I could have one section graded based upon individual pre-lab homework with re-done individual homework after the lab. The other section could have individual pre-lab homework and then group homework due at the end of the lab. Comparing individual improvements using individual homework problems and related problems on exams could provide evidence for whether or not group work helps the individual improve performance and if this is an effective means of teaching. Comparing the individual pre-lab homework to group homework is another aspect of the research that would help prove or disprove the existence of the Assembly Bonus Effect and answer the question in my class about whether or not groups really do come up with better answers together compared to what any one individual could create.
Effective Teamwork

The next issue I need to address in CH459 is the issue of the free-rider, the lone cadet that I find twiddling thumbs or staring into space while peers toil away. How can the course be designed and teams set up to maximize performance and learning? For this problem, I turn to economists, game-theorists, and experienced educators.

Grading and Game Theory

In an article titled “Optimal Incentives for Teams,” Yeon-Koo and Yoo compare relative performance evaluation (RPE) to joint performance evaluation (JPE). [6] Essentially, RPE means that everyone works as an individual competing for scarce resources. There are only so many promotions and so many bonuses to award and each individual is fighting against the other to succeed.

In JPE, teams compete against other teams. If the team does well, everyone in the team does well. This encourages cooperation and peer pressure pushes each member to contribute utmost effort. Yeon-Koo and Yoo present various examples of where JPE helped corporations perform better, using examples that range from AlliedSignal Aerospace increasing revenue by 11-percent after switching from RPE to JPE to Eastman Kodak decreasing production defect levels by 70-percent when switching to team evaluations. [6] Based on these examples, Yeon-Koo and Yoo develop mathematical models to simulate wage incentives and performance levels and reach the conclusion that the optimal incentive system rewards employees when members of their team do well and when work is done together, not separately without others knowing how much work each other member is contributing so that effective peer-sanctioning can occur. [6]

The researchers’ model assumed that workers can either shirk or work. Making the choice of effort level a binary option simplifies the math model (although it is still not so simple to look at for the untrained eye). When a JPE system is in place and peer sanctioning is possible (reducing wages for peers or even firing them), a Nash equilibrium is reached (the case where the optimal good for the individual and the group exists) when every member of the team works.

The analogy in the classroom for RPE grading would be grading on a curve. It would be in each student’s interest to encourage the others not to study and to perform badly so that their individual performance will be better rewarded. This study would suggest that this sort of grading would not produce the best outcomes, and fortunately I do not employ this method in class assessments.

A much earlier study investigated incentives in the workplace in the early 1950’s and reached similar conclusions, that JPE compensation was much more effective than RPE, even when incentives for collaboration and supporting other workers were incorporated into individual compensation. [7]
The example of JPE breaks down in the classroom for CH459 because there is currently no real mechanism for peer sanctioning beyond writing up a negative peer evaluation that does not affect the final grade. Poor students cannot reasonably be fired from the team in a lab setting.

Without peer-sanctioning, does JPE (i.e. a group grade) still work though? For some insights about this I turn to other examples from educators.

**Student Views on Group Work in Higher Education**

The student that complained to me about group work had two main complaints: 1) he ended up doing the lion’s share of the work while others watched, and 2) he did not learn much from his peers. To expand upon this anecdotal evidence concerning student satisfaction with group work, surveys of larger student populations provide more statistically significant conclusions about student views on group work.

A team of Spanish professors who teach business administration and management decided to assess student satisfaction with teamwork after a group project and their conclusions offer some interesting insights that apply to CH459. [8] Simply put, the educators used surveys to determine how much time each team member invested, how useful students estimated the assignment was in enhancing their knowledge, and how satisfied the students were with group work. Based on student responses, the study authors concluded that, “group work is a good teaching and learning tool,” but also found that “there are students who shirk their responsibilities fully.” [8] Despite the free-loaders, the majority of students stated that group work increased their knowledge base and reported being happy with doing group work.

Another questionnaire based article supports this general conclusion. Joseph McIntyre of Auburn University was interested in determining if group work increased freshmen motivation in an introductory engineering course. [9] McIntyre found that students felt group interactions were more engaging than standard lectures and increased student interest in the course. [9] An ancillary note in the study that applies particularly to CH459 is that students especially enjoyed group work when the questions being asked were open-ended without a definite right or wrong answer. This applies to CH459 because cadets are given the opportunity to design their own experiments and establish their own procedures. There are many ways to solve the problems and no definite right or wrong way. Keeping this aspect of my course, or even expanding it so that there are more creative portions of the group project may enhance student enjoyment of group work and satisfaction with the course, which would presumably motivate cadets to want to learn more and accomplish more during lab periods.

**Collaboration and Performance**

One group of educators conducted research on collaborative and non-collaborative teams. [5] In the collaborative group, team members worked together and all received the same
compensation (small amounts of money in the study). In the non-collaborative group, team members knew that they were part of a team and that they would all receive the same compensation, but they did not know who the other members of the team were and could not see the other members working or know what the others answered. They knew that the best answer from the group would be compensated and that everyone in the team would receive the same compensation amount. The non-collaborative group members worked as individuals, but knew they were teammates.

The researchers found that the non-collaborative group members answered questions correctly less often than collaborative groups and engaged in what the researchers deemed “social loafing,” assuming that someone else in the group would answer the problem correctly and therefore not putting in as much effort to solve the problem. [5]

While this is clearly an extreme example, it shows the importance of teammate interactions. This is an issue that I could investigate with the pre-lab and post-lab homework study. If cadets believe that the pre-lab homework does not count for as many points as the post-lab homework that is done as a group, they may participate in “social loafing” and wait for the group to provide answers for the post-lab.

I think the other point that this study raises is that cadets in a team must be afforded time to actually collaborate. If the lab setting is too harried and not enough time is given, cadets will be forced to break the assignment into parts and work as individuals who then assemble the work together at the end. While this approach has the advantage over a pure non-collaborative team in that the cadets know their individual work will be graded, it shares the same disadvantage of aggregate knowledge not being shared, and the team does not perform to the highest potential.

**Free-Riding, Student Attitudes, and the Value of Teams**

Another pair of educators investigating group projects caught my eye just with the title, “Does It Matter if I Hate Teamwork? What Impacts Student Attitudes toward Teamwork.” [10] If any of the studies that I looked into might help me address my student’s complaint, it ought to be this one. The essential problem that my student was expressing was the free-rider issue.

What Pfaff and Huddleston conclude is that project grades, perceived workload, time allotted in class, use of peer evaluations, and absence of a free-rider problem are all critical indicators of how satisfied students are with group work. Not surprisingly if grades were good, people tend to be happy about the outcome of the group work. When the task was doable in the time given during class, students were more satisfied. This likely relates to the work done by Besedes et al. [5] because when students are outside of class they are more likely to just work as individuals as opposed to coordinating their busy and conflicting schedules to meet outside of class. Peer evaluations can serve as a way to hold each team member accountable (a form or peer sanctioning potentially).
The most useful aspect of this study is the conclusion reached about how to avoid free-riders: allow members to change groups. [10] Reading this article, I nearly yelled out, “Eureka!” Allowing members to self-assemble would basically allow students to expel non-performers the way a private company can fire deadbeats. Cadets would try to do their best to stay in a group with high performers to get good grades.

Conclusions

Despite the conflicting views on whether or not groups outperform highest performing individuals, there are definite and compelling reasons for using a group learning approach in a course like CH459. As Pfaff and Huddleston point out, there are real, practical reasons for putting cadets into groups, especially in a laboratory environment. An important reason that they mention is maximizing student/faculty interaction. [10] It is hard, if not impossible to interact meaningfully with many students on a one-to-one basis in a lab. In such an environment, being engrossed in conversation with one cadet while all the others are left to their own devices could actually be dangerous. Breaking the class into groups allows me as the instructor to more easily ensure that each group gets access to me and that I can help make sure equipment is operated safely.

While group work is necessary for simply pragmatic reasons, it may also offer an improved learning environment. Earlier it may seem like I was just finding the silver lining in my student’s complaint, but I truly do believe that in almost every human endeavor, even the most apparently reclusive, we are all in the people business. Cadets really do need to learn how to effectively work in teams and how to motivate others to perform. These are key abilities for an officer.

As the course director, I have an obligation to assist with ensuring all cadets have the proper motivation in class and that assessments are fair and create incentives that best foster learning. Based on the literature review I have conducted, I believe there are a few simple changes I could make to CH459 that would have radical effects.

The easiest to implement and likely most effective measure will be to allow cadets to change groups. I will assign groups with a mix of GPAs for the first lab like I have before, but I will make it known that the next lab cadets will pick their own groups. This might motivate each cadet to work well and contribute or face losing high performing teammates in the next lab experiment iteration.

The other aspect that is lacking in CH459 that would create the ideal JPE environment according to the work of Yeon-Koo and Yoo is peer-sanctioning. [6] Peer evaluations could be used to determine grades to some small extent. Cadets could complete a survey rating all of their teammates and this could translate into part of an instructor grade. I would make this a minute percentage of the overall course grade, but cadets in chemical engineering are often concerned
about small point values that are nearly insignificant in comparison to the total number of points in the course, so peer-evaluation grading could be quite effective.

**Future Inquiry**

For safety reasons and in order to be able to handle the sheer volume of grading that must be accomplished even with group work in CH459, I will maintain groups in the future, but that does not mean I cannot experiment with the course and make adjustments to discover improvements.

One advantage that I have is that there are two sections, which makes for a convenient way to establish a test and control group. The small sample size makes reaching statistically significant conclusions difficult without many repetitions, which is hard to achieve during one tour as a rotating instructor, but research in the classroom might at least give me some intuition about best practices for laboratory courses.

I am definitely curious about the Assembly Bonus Effect and if cadets working together actually outperform highly proficient cadets working alone. Depending on program director guidance, I may alter some group and individual graded assignments and compare individual results on exams, along with homework performance. I would be curious to see if cadets do better when they complete the homework alone before the lab, receive feedback during the experiment, and then attempt the homework again or if they do better when they do the homework together on the second iteration. This individual versus group comparison could also be made for designing experiments. While it may be hard to quantify, it is possible to identify procedures that more elegantly test a hypothesis than others. It would be interesting to compare the experimental designs that individual cadets create on their own versus what they can generate as a group.

**Final Thoughts**

It is useful to reflect upon student feedback; even anecdotal evidence can be helpful for improving a course. That said, before changing my course, I want to have more solid evidence, and that is what this literature review is useful for. Educators have been facing issues with group work far longer than I have been an instructor. Employing scientific research about best teaching practices will help me improve my course and avoid turning my cadets into guinea pigs whose learning or cadet careers might suffer if I make poor decisions as an educator. The simple step of allowing cadets to pick their own groups is something that I had not previously considered, and I believe it will be a big improvement to the course.
Annotated Bibliography


This book challenges the common ideal of the extrovert leader and presents the argument that introverts working in quiet office spaces, often alone for long periods, may produce better outcomes than teams do in business, design, and generally any area where deep thinking is necessary, which is most areas of modern life.


Huff’s book was recommended to me in graduate school by a professor of statistical mechanics as a classic handbook on how statistics are often misused and leave people confused or misled.


This study found that groups perform better than the highest performing members of the team do alone and claims to be proof of the Assembly Bonus Effect. A sample size of 222 groups consisting of 3-8 members each were given individual questions and then worked in groups to answer the same questions. Overall scores were higher for group work than individual work.


Tindale and Larson argue that the statistical method of analysis in the work by Michaelsen *et al.* makes the conclusions invalid. This study concludes that the Assembly Bonus Effect has not been proven to exist and that groups do not perform as well as an individual member on an isolated question. Only in aggregate scores do group averages surpass individuals.


In this research, some teams are allowed to work together and other teams work individually, yet the whole team receives credit for the best answer from any of the individual work submissions. These non-collaborative teams performed poorly and members were more likely to free-ride and expect others to do the work.
This research highlights examples of corporations that improved overall performance by rewarding team efforts instead of individual efforts. They term these two methods Joint Performance Evaluation (JPE) and Relative Performance Evaluation (RPE) respectively and found that JPE incentives produced better outcomes for the organization as a whole.

Similar to the work of Yeon-Koo and Yoo, this study discusses workplace incentives and finds that team rewards lead to more collaboration and better results than individual rewards. Even rewards to individuals for demonstrating teamwork were not effective in producing outcomes equivalent to the results with cooperative group incentives.

This team of Spanish professors surveyed their college students regarding their opinion of group work and students overwhelmingly had positive responses, even though there were students who contributed nothing to the group and were free-riders.

This professor surveyed students and found that they enjoyed group work, especially when discussing open-ended questions more than normal lecture.

These educators emphasize the practical benefits of group work, such as improving the feasibility of faculty/student interaction, and they offer recommendations about how to make group work more effective. One example is helping to reduce free-loading by allowing students to pick their own groups after each iteration of an assignment when group work is a repeated event.
Additional References


