

Group formation on physics exams

Steven F. Wolf,¹ Cody Blakeney,² and Hunter G. Close²

¹*Department of Physics, East Carolina University, Greenville, NC 27858*

²*Department of Physics, Texas State University, San Marcos, TX 78666*

As our classrooms become more active and collaborative, we need to consider ways that our assessments can take on the same active and collaborative spirit that our classes have. One way that we can accomplish this goal is through the use of group exams. In exams at Texas State University and East Carolina University, we have utilized a two-phase exam format allowing open collaboration—that is, allowing students to form their own groups. We have analyzed self-reported collaboration data and find that the room influences how students will associate with each other. We have also compared collaboration patterns evident in student responses for students in open collaboration settings to students in closed collaboration settings.

I. INTRODUCTION

Modern science is built on the foundation of collaboration—working together to design experiments, communicating results, and using those results to create or refine scientific models—making it a key combination of several scientific practices identified by the NRC[1]. Therefore students must not simply learn new ideas and how to apply them in practice, but also how to work together in groups and engage in productive scientific discourse. In response to the many national calls promoting active learning (e.g. [1]), and in order to make teaching practice mirror scientific practice; more instructors are using group-based activities in their classrooms.

One of the ways that teachers can communicate the importance of group work is to assess students in group settings. To that end, group exams have come into use in STEM classrooms across multiple disciplines[2–6]. While each implementation is slightly different, a common implementation is the two-phase exam. The two-phase exam is when an exam is given twice, once while students work in a traditional exam setting on their own, and then again while students are allowed to collaborate. In some instances, the exam is identical, while in others, the exams are similar, but not identical.

At East Carolina University and Texas State University, some instructors have been giving two-phase exams in our introductory calculus-based physics courses. Our implementation (detailed in the next section) is similar to the one used by Beatty[2], with one key difference. We allow open collaboration among all students during the exam period. This difference allows us an insight into our students that is not present when an instructor fixes group membership. In particular, we are able to see how students organize themselves in class, explore the different ways that the classroom can influence this collaboration, and begin to gain some insight on how students might be choosing collaborators. We will use tools developed in the network analysis community to describe the networks our students create on group exams. In the next sections, we will discuss the details of our implementation of group exams, look at different ways that the classroom structure influences student collaboration, and examine some of the different collaboration metrics introduced by Beatty[2] in

our classes.

II. IMPLEMENTATION

In our implementation, we have chosen to use duplicate exams in the individual and collaborative phases *for all exams in the course*. For courses at both institutions, the class period lasts for 75 minutes. Instructor A used a multiple choice format and had students take the individual exam in (approximately) the first half of the period. Students were then immediately given the group exam, keeping their question sheets so that they could have their exam responses for the individual portion. Instructor B used a hybrid multiple choice/free response exam for all mid-term exams, and a multiple choice exam for the final. For the midterm exams in instructor B's course, the students were given the full time to complete the individual exam. The group exam was given during a later class period or recitation section, but students did not have access to their individual exam answers. For the final exam, the two and a half hour exam period is split in half (or slightly more time is given to the individual phase), and the group exam follows immediately. In order to encourage full participation in the group exam, students are given exam credit for their work on the group exam. Typically, they are able to gain up to half of the points that they missed on the individual exam without penalty. This may lead to some game-playing on the group exam—students may decide to change their answer to “cover their bases” or for reasons other than they think that the marked answer is correct. For the duration of this paper, we assume that students are answering questions on the group exam because they think that it is the correct answer, and not for any other reason.

In both the individual and group exams, each student has a copy of the exam, and turns in an individual answer sheet. However, we have chosen to allow open collaboration on our exams. This is for two reasons: First, this is how we have chosen to run our own active, collaborative classrooms, and we do not wish to change the classroom environment for examination purposes. Second, we feel that this is a way that we can get some data about the class's social network outside of class, and we hope to leverage it for further study. Research

TABLE I. Summary of exam data. The number of exams includes the final, and the number of students is the number of people who took the final exam. Both instructors have a PER training. Instructor A and B were at different institutions.

Semester	Instructor	Num Exams	Course	Num Students
Fall 2014	A	5	Heat & Waves	47
Fall 2015	B	4	Mechanics	44
Spring 2016	B	4	E&M	36

utilizing network analysis focused on the development of student social networks is ongoing in the broader PER community (e.g., Brewe et al.[7]). In particular, at Texas State University, the physics department has been paying close attention to how the campus physics community can be enriched by focusing on the introductory physics courses[8]. We collect the following data from these exams:

- Individual-phase multiple-choice responses
- Group-phase multiple-choice responses
- Responses to “On this exam I *mostly* worked with...” and “On this exam I *sometimes* worked with...”

All analysis described in the sections below involves only the multiple choice portion of the exams. Table I summarizes the data streams reported on here.

III. CLASSES AS SOCIAL NETWORKS

An active, collaborative classroom is a natural setting for student’s social connections to manifest themselves. The PER community is familiar with the use of graph theory or network analysis techniques being applied to different kinds of relational data (e.g. [7, 9]). In particular, Brewe et al. applied network analysis techniques to students working in a physics learning center.[7] Our method for parameterizing social networks from self-reported data is drawn on the method Brewe et al. developed. In order to draw a network graph, we utilized student responses to the first prompt: “On this exam I *mostly* worked with...”. A line (edge) connecting two people (nodes) was drawn between person A and person B if A reported working with B, if B reported working with A, or if both A and B reported working with each other.

Two different classrooms were used in the study. In Fall 2014, instructor A taught in a terraced classroom with movable desks. The terraces make the construction of large groups difficult, since many desks cannot be moved close enough together to promote meaningful large-group collaboration. During class and on exams, portable white boards, markers, and erasers were made available to students as a shared work space. In Fall 2015 and Spring 2016, instructor B taught in a flat classroom with large round tables seating 8 around. During class and on exams, white boards—mounted to the interior walls of the classroom, markers, and erasers were available to students as a shared work space. All classes utilized online and written homework, tutorial-style curricula,

and offered out-of-class screencasts supporting a “flipped” pedagogy. Given the high number of similarities between the courses, differences in the collaboration practices may be due to the classroom design.

A key aspect of the group learning literature base is the focus on the group as a basic unit. Allowing open collaboration flies in the face of that since students can freely flow around the classroom and talk to any other student that they wish. So in order to compare group work in environments with open collaboration to static group environments, we will need to re-create groups as best as possible. Fortunately, network analysis has tools to allow us to solve this problem. All analyses described here have been carried out using the open-source statistical software R[11]. The `igraph` package has been utilized to carry out analyses of the network structures[12]. In order to explore potential effects of the physical classroom on student collaboration, we will visualize the networks, and compare their network properties, and in order to reconstruct groups, we will utilize community detection algorithms developed by the statistical physics and applied mathematics communities[13]. In particular, we re-created groups using the `cluster_edge_betweenness` algorithm for the following reasons. First, Newman and Girvan recommend this algorithm for networks with smaller numbers of nodes (e.g., less than 100). And second, when the algorithms that we tested gave different output, they tended to produce even larger groups, encompassing up to half of the class. In fact, no student reported working with more than 9 classmates either “mostly” or “sometimes”.

Our graphs (see Fig. 1) allow us to note a few differences between the classes. First, notice that the groups tended to be larger in Fall 2015 and Spring 2016. This makes sense due to the large tables in that classroom. This is borne out in the numerical data shown in Table II. Fall 2015 saw groups that, on average, had three more members than groups in Fall 2014. And the largest groups in Fall 2015 were often twice as large as the largest groups in Fall 2014.

Next, we notice a trend across exams—we notice that the number of groups decreases, and the size of groups increases. We will also find it instructive to look at the median edge density for each of the algorithmically identified communities. A community’s edge density is defined to be:

$$\text{edge density} = \frac{\# \text{ edges drawn within a community}}{\# \text{ possible edges in a community}} \quad (1)$$

Since the group size is growing throughout the semester, and the edge density is (generally) growing or remaining constant, we have evidence that the students are becoming more connected throughout the course of the semester. This is evidence that students are building a larger community within the scope of the class.

Lastly, a chief concern that we get when we tell other educators that we let students pick their own groups is, “The smart students will just work together,” (and leave the other students out). Our data tell a different story. Each node in Fig. 1 is colored by their score on the individual exam (grouped

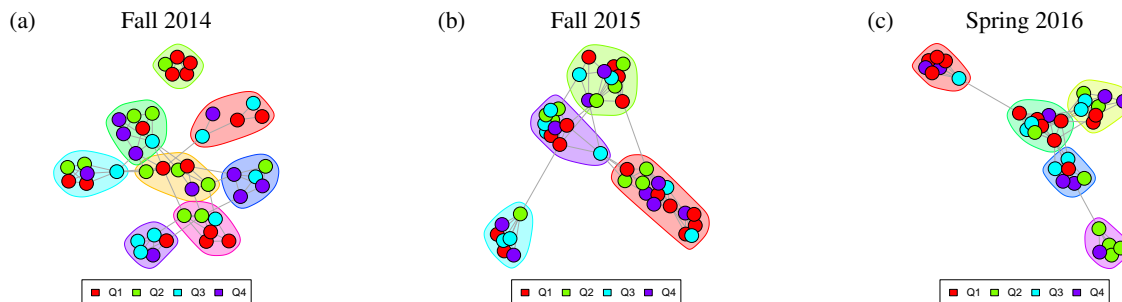


FIG. 1. (Color online) Representation of collaboration on a group exam. On each graph, a node is a student, an edge is drawn between collaborators, and communities are encircled and shaded in. Nodes in the graph are laid out using the Fruchterman-Reingold algorithm.[10] The coloring of each node represents the quartile that each student scored in for the individual phase of the exam (Q1 is the top quartile). The color of the group bubbles is for identification of the groups and has no further meaning. The networks shown are from the final exam of each semester. (a) is from Fall 2014, instructor A, (b) is from Fall 2015, instructor B, and (c) is from Spring 2016, instructor B.

by quartile). Based on this, and an inspection of these graphs from other exams (not shown here), we do not see systematic evidence that students are trying to corner the intelligence market while choosing their collaborators.

TABLE II. Group properties for all exams in study.

Exam	Num. of Groups	Max Group	Avg Gp Size	Med Edge Density
Fall 14 EX 1	17	6	2.706	0.667
Fall 14 EX 2	8	8	5.875	0.967
Fall 14 EX 3	8	7	5.875	0.757
Fall 14 EX 4	8	7	5.625	0.833
Fall 14 EX 5	8	7	5.375	0.817
Average	9.8	7	5.091	0.808
Fall 15 EX 1	6	13	8.167	0.948
Fall 15 EX 2	6	14	7.500	0.528
Fall 15 EX 3	7	12	6.286	0.952
Fall 15 EX 4	4	15	11.000	0.718
Average	5.75	13.5	8.238	0.787
Spr 16 EX 1	6	8	6.000	0.900
Spr 16 EX 2	6	9	5.833	1.000
Spr 16 EX 3	5	10	7.000	1.000
Spr 16 EX 4	5	9	7.000	0.944
Average	5.5	9	6.458	0.961

Another way that we can suggest an equivalence between open collaboration and closed collaboration is by comparing these students' exam performance with students' exam performance while working in fixed groups. In 2015, Beatty came up with five metrics for discussing collaboration on group exams.[2] We apply each metric to our data, using the groups identified by the community detection analysis. We will only report the final exam for each class in our data set, as that is the exam that was reported on by Beatty. Fig. 2 and Fig. 3 display our results and compare them to Beatty's on a

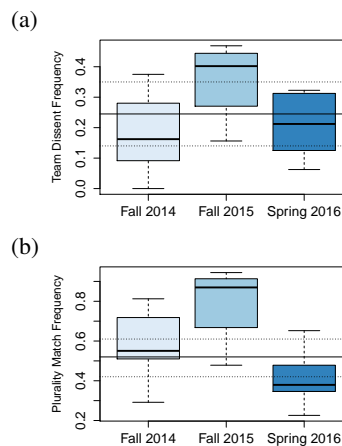


FIG. 2. (Color online) Comparison of our final exam group data with Beatty's data.[2] Solid horizontal lines indicate the median value reported for all of Beatty's exams and the dashed lines indicate an estimated inter-quartile range for the following measures: (a) Dissent Frequency and (b) Plurality Match Frequency.

basic (non-statistical) level. These measures show some variability in collaboration practice, suggesting that these measures may be valuable for characterizing the collaboration in groups in STEM classrooms.

Dissent Frequency: Describes how often students who disagreed initially, continued to disagree on the group exam.

Plurality Match Frequency: Describes how often the group picks the most popular answer from the individual exam within the group (as long as the most popular individual exam answer was not unanimous).

Leadership Match Frequency: The maximum match frequency for each group. In order to calculate the match frequency, each group member's individual exam responses is compared to the group's consensus responses.

Answer Changing Frequency: The minimum fraction of answers that was changed from the individual to group exam by a single member in a group.

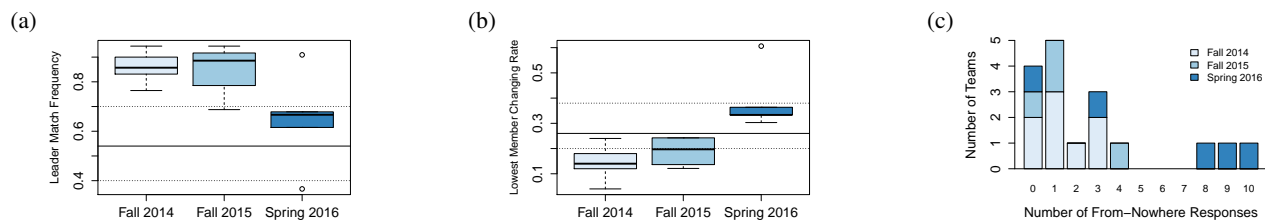


FIG. 3. (Color online) (a) Comparison of our final exam group data with Beatty’s data for how often groups simply “followed the leader.” (b) Comparison of our final exam group data with Beatty’s data for answer changing frequency. (c) From nowhere response distribution. The Fall 2014 and Fall 2015 semesters were in line with Beatty’s distribution, however the Spring 2016 class had noticeably more consensus responses not on any individual group member’s exam.

From nowhere responses: Consensus responses on the group exam that no group members chose during the individual exam.

The Leadership Match Frequency seems to be high for the Fall 2014 and Fall 2015 classes. This could be a consequence of allowing students to pick their own groups (they choose to work with someone that they trust to get the answer right). The other difference we saw is in the high number of “from nowhere” responses in Spring 2016, which is the subject of future study in our group. Otherwise, the measures exhibit the variability seen in the closed collaboration data.

IV. CONCLUSION

Collaboration is an important part of modern science, and in order to emphasize the importance of collaboration, many instructors are using group exams. We used a special kind of group exam, called a two-phase exam, and allowed students to choose their own collaborators. We studied student collaboration networks using the mathematical framework of network analysis. In particular, we found that the classroom environment had an impact on how students formed groups. This was seen both by visual inspection of the network graphs (Fig. 1), as well as by looking at the size of groups and the

density of connections within groups. We also find little evidence suggesting that students segregate themselves into performance castes based on our observation of the networks.

Using these network analysis tools, we were able to reconstruct groups, so that students working in open collaboration environments can be compared to static groups. We compared our students collaboration characteristics to those from Beatty’s study using the measures developed there.[2] Comparisons between open collaboration environments and closed collaboration environments do not show major differences between the different types of collaboration. However, further study is required to determine what meaningful differences are for these collaboration parameters. From the perspective of group exams, there does not at this time seem to be a strong reason to prefer one form of collaboration to another.

In the future, we plan to probe student thoughts about group exams to better understand how the measures we have defined should be interpreted, and to propose additional measures. Also, given this new data set (student interaction data) we can explore questions such as, “Who is a team player?” or “Who is isolating themselves?” and consider different ways that the answers to these questions can improve our instruction, as well as improve our ability to include as many students as possible into the physics community.

[1] S. R. Singer, N. R. Nielsen, and H. A. Schweingruber, *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering* (National Academies Press, 2012).

[2] I. Beatty, in *Physics Education Research Conference 2015* (College Park, MD, 2015), PER Conference, pp. 59–62.

[3] Y. Lin and D. Brookes, *AIP Conference Proceedings* **1513**, 254 (2013).

[4] C. E. Wieman, G. W. Rieger, and C. E. Heiner, *The Physics Teacher* **52** (2014).

[5] B. H. Gilley and B. Clarkston, *Journal of College Science Teaching* **43**, 83 (2014), ISSN 0047231X.

[6] H. Leight, C. Saunders, R. Calkins, and M. Withers, *Cell Biology Education* **11**, 392 (2012), ISSN 1931-7913.

[7] E. Brewster, L. Kramer, and V. Sawtelle, *Physical Review Special Topics - Physics Education Research* **8**, 010101 (2012).

[8] E. W. Close, J. Conn, and H. G. Close, *Phys. Rev. Phys. Educ. Res.* **12**, 010109 (2016).

[9] S. F. Wolf, D. P. Dougherty, and G. Kortemeyer, *Physical Review Special Topics - Physics Education Research* **8**, 010124 (2012).

[10] T. M. Fruchterman and E. M. Reingold, *Software: Practice and experience* **21**, 1129 (1991).

[11] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2015), URL <https://www.R-project.org/>.

[12] G. Csardi and T. Nepusz, *InterJournal Complex Systems*, 1695 (2006), URL <http://igraph.org>.

[13] M. E. J. Newman and M. Girvan, *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* **69**, 1 (2004), 0308217.