Student Cognition in Physics Group Exams

Timothy Sault,[1] Hunter G. Close,[2] and Steven F. Wolf[1]

[1]*Physics Department, East Carolina University, E 5th Street, Greenville, North Carolina, 27858 USA*
[2]*Physics Department, Texas State University, 601 University Drive, San Marcos, Texas, 78666 USA*

Exams are intended to give faculty information about what students know, and where they need more support. But traditional analyses that get shipped with the exam results received from the university scoring office focus on the exam items, not students' ideas. Moreover, one of our goals for students is building their cognitive network of physics concepts. With sets of contextually related questions, we can analyze students' cognitive network as applied to a particular context. We report on the effectiveness of existing tools in identifying patterns in student responses on both individual and group exams.

I. INTRODUCTION

When we choose to create active and collaborative environments for our students, we are sending a message both about how they learn best, and also about the importance of being able to work collaboratively in STEM fields. One of the ways that we can underscore the importance of this message is by using active and collaborative assessments. One type of active and collaborative assessment is the group exam. Group exams have been in use for some time now, and are gaining popularity in physics classrooms. PER scholars have also been studying how physics exams promote collaborative learning environments [1, 2], how group exams help us assess student learning and transfer [3, 4], and the ways that collaborative testing affects performance and information retention [5]. Group exams are an opportunity for students to get "peer review" on their responses from the individual exam. One such implementation of group exams is the two-phase duplicate exam format, where the first stage is an individual exam taken in a traditional fashion, and the second stage in which students may work with their in class group as well as any other student in the classroom. In our research group, we have been studying group exams in open collaboration settings [6, 7] using tools of network analysis.

Network analysis encompasses a rich set of tools that have been gaining popularity in PER [8], that are being applied broadly. Some studies have focused on studying cognitive networks [9, 10]. Others have used social networks to predict academic performance [11] and analyze the structure of multiple choice exams [12]. Still others are studying PER collaboration networks [13]. Our group has been studying classroom social networks in the context of group exams.

The classroom environment and exam structure of a collaborative classroom can significantly change how students form groups in these exams. Previously we saw that the design of a classroom had a significant impact on the size and structure of classroom networks [6]. We have also learned that these student groups expand in size over a semester and the classroom as a whole becomes more densely connected as students get to know more of their peers [6]. One of the properties of social networks is that relative status can be inferred from the directionality of ties within the network [14]. We found that as the semester goes on, classroom networks take on this property of social networks; students are able to identify peers who can help them, and preferentially select these students as collaborators [7].

However, we do not know how group exams change the ideas that students express about physics concepts. Previous work [6, 15] has focused on answering questions like "*How often do students change their answers from individual exams to group exams?*" and "*Do students simply copy the answers of the 'smartest' student in their group?*", but has not examined common non-normative ideas and the ways that group exams change student conceptions about physics. Given the fact that students score so highly on group exams (often 95% or greater), the wrong answers students give during the group exam are indicative of physics ideas that students *do* hold stably when confronted with different points of view through collaboration.

Although non-normative ideas students present upon retesting in group settings do suggest that they are more deeply held, student concepts of physics are dependent on context [16]. With this in mind, in order to probe how students connect different physical concepts, we build exams by presenting a few single contexts and asking sets of questions probing multiple concepts. Students are prompted to think about each piece of their conceptual model and whether they logically fit together. We believe working in a group environment may help students better identify and solve these contradictions in their conceptions of physics.

In this paper, we examine the effectiveness of cognitive networks in analyzing common misconceptions on individual exams, and consider how they change on group exams. Brewe *et al.* used cognitive networks to analyze data from the Force Concept Inventory; an exhaustively studied assessment for which there is much data [12]. Given that typical exams are not as rigorous as research validated assessments, and typical class sizes are small (so that research-grade validation of an exam is impossible) we will employ item-specific tools of Classical Test Theory (CTT) [17] to aid in the interpretation of our results. In Section II we will outline how MAMCR and our CTT measures work, and in Section III we will apply these techniques to a particular exam with the following research questions in mind:

- Can cognitive networks be used effectively to identify student thinking patterns in small tests which are not well studied?

- What ways are congitive networks changed by group exam collaboration?

II. METHOD

Despite knowing what we want from our exams; writing exams which reliably assess what we want them to assess can be difficult. Interpretation of exam results often ends at submitting grades. By using a combination of item-specific methods from CTT [17], the module analysis for multiple choice responses (MAMCR) from Brewe *et al.* [12], and answer chain analysis, we will examine the cognitive networks of small test data and how they are affected by group exam collaboration.

Classical test theory is a set of measurements which test the overall and item-specific reliability of tests and test items [17]. Global test measures from CTT recommend a minimum student number of 100 students [17]. Our smaller class size means we will focus on item-specific measures. On the other hand, MAMCR is an intuitive method of observing underlying physics concepts by mapping item responses in a network where vertex size is dependent on answer frequency and tie weight is determined by how many students chose both responses [12]. Answer chain analysis simply looks at which sets of answers students chose for a subset of questions and attempts to assign a conceptual model to each common chain.

Classical test theory provides two item-specific measures useful in this case [17]:

Item Difficulty: The fraction of students who answered an item correctly. (Acceptable range: 18-85%)

Item Discrimination: Measures how well an item discriminates between those who do and do not understand the material. It is calculated as the difference in item difficulty between the top-scoring half of the class and the bottom half. (Acceptable range: >20%)

With the acceptable difficulty and discrimination thresholds, we can use question fitness for network context which is more related to test design.

Answer chain analysis is a quick method of extracting student response patterns from test data that is especially useful for studying sets of questions related by the same context, such as the set described in Figure 1. The idea behind creating question sets like this is to mirror how you might evaluate a multi-part free-response question. After identifying and measuring the frequency of unique response patterns to sets of questions, we can qualitatively analyze each ‘chain’ by connecting them to conceptual ideas. However, gleaning patterns from these sorts of counting statistics becomes more difficult as the number of questions and/or the number of choices increases.

Fortunately network analysis tools can be applied to help us understand these patterns through visual and algorithmic processes. To convert our exam responses into networks, we utilize the MAMCR procedure developed by Brewe *et al.* [12]. Each possible response is visualized as a vertex; colored by a CTT color scheme, and vertex size determined by the frequency with which students chose them. Vertices are connected by edges; the thickness of which indicates how many students chose *both* responses. The vertices and edges are then placed optimally by the Fruchterman Reingold layout algorithm. Due to their overwhelming influence on the network structure, correct answers are removed from the network [12]. The network is then divided using a clustering algorithm and analyzed as conceptual subsets [12].

Using knowledge of which items are contextually related question sets and which responses are in the same chain, we can focus our analysis of the network clusters. Alongside qualitative analysis of the responses and their conceptual consequences, classical test theory allows us to determine the significance of connections and structure in these networks. While MAMCR analysis removes correct answers from the network, answer chain analysis helps us flesh out misconceptions by observing how the incorrect answers can be conceptually related to correct ones by students.

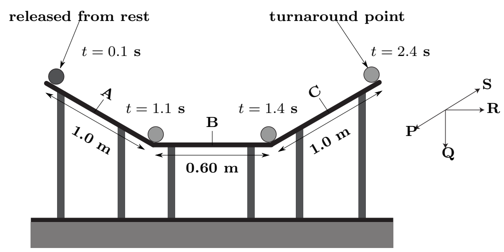
When looking at an exam with MAMCR and CTT, we first start by clustering and labeling the nodes according to our criteria. We examine each response cluster individually by first trying to identify conceptual connections between the items, then determining if the connections are caused by test item reliability by using our CTT methods. This conceptual analysis is best guided by the results of our answer chain analysis; as items from the same chain are much more likely to be conceptually related by test design.

The main set of contextually related questions from this exam, which contained the most identifiable physics misconceptions, was focused on a ball rolling on a double ramp shown in Figure 1. Students were asked to determine speed, magnitude of acceleration, and direction of acceleration at multiple points on the ramp. Along with the correct answers, distractors related to common misconceptions and common mathematical errors were present in the multiple choice responses. This context and the diagram referenced therein, are shown in Figure 1.

III. RESULTS

For the purposes of this paper, we will be examining the first exam from our fall 2015 university physics class in which there were 44 students that responded to 10 multiple choice questions and a free-response section which is not analyzed here. This set of students took both the individual and the group exam. This exam featured a network which contained several nontrivial clusters, and consisted mostly of ‘good’ questions.

The main individual exam MAMCR network for the first exam in Fall 2015 (Figure 2(a)) features a dense central structure and a few peripheral clusters. The clear cluster will be excluded, because we could not determine conceptual significance from the three-response cluster. The individual data



Questions related to this scenario (numbered by exam question number):

4. Determine the speed of the ball at $t = 1.1\,\mathrm{s}$.
5. Determine the *magnitude* of the acceleration of the ball at point A (Halfway up the first incline).
6. Indicate the *direction* of the acceleration of the ball at point C (Halfway up the second incline).
7. Determine the *magnitude* of the acceleration at the turnaround point $t = 2.4\,\mathrm{s}$.

FIG. 1. Double ramp context from Fall 2015 Test 1, based on a problem created by the authors of the *Tutorials in Introductory Physics*. This scenario was the main point of focus for student misconceptions on this test; However, there were 10 total questions on the assessment.

network mostly consisted of 'good' questions which were both discriminatory and difficult. While there were no 'bad' questions, there were a significant number of questions which were sufficiently difficult, but not discriminatory. This indicates that these questions were equally difficult for the whole class, regardless of overall score.

The central yellow cluster (Figure 2(b)) from the Fall 2015 network related to a problem where a magnet was holding paper to a refrigerator. The choices in this cluster were indicative of a misconception that static friction was always proportional to normal force (Q3A and Q1D). While these responses were the most well connected in this cluster, this and other misconceptions from this context, did not show up prominently in other clusters. This indicates that the misconception was isolated from other physics misconceptions.

The dark green cluster (Figure 2(c)) from the Fall 2015 network is a small but conceptually significant cluster from the double ramp context. It represents the idea that acceleration is $9.8\,\mathrm{m/s^2}$ while rolling down one of the ramps. One of the logical conclusions of this thought process is that the acceleration should be straight down, as the ball is accelerating exactly as it would in free fall. However, this response (Q6B) is located the yellow cluster (b). Using the answer chain analysis, we discovered that the responses in this network were more commonly associated with an acceleration direction down the ramp; the correct answer.

The largest group from the Fall 2015 network, the brown cluster (Figure 2(d)), represents the other main misconception from the double ramp problem; improperly recognizing the conceptual difference between acceleration and velocity. For question 4, choice B is calculated using $v = \frac{\Delta x}{\Delta t}$ for the

Misconception	Average response frequency Individual	Group
$F^{f_s} = \mu_s F^N$ always	22.5	2.5
$\lvert\vec{a}\rvert = 9.8\,\mathrm{m/s^2}$ always	13.3	5.0
Confuse $\vec{a}$ and $\vec{v}$	16.25	3.75

TABLE I. Average response frequency for each of the choices identified on the individual exam and how those change for the group exam. This is calculated by the total number of responses in each misconception group / the total number of nodes in the misconception group.

portion of the motion down the ramp. In question 5, choice A uses the velocity indicated by 4B to calculate the acceleration of the ball as it comes down the ramp using $a = \frac{\Delta v}{\Delta t}$. which should be zero based on the assumption they implicitly made to answer question 4. Students then chose in question 6 that the *acceleration* points up the ramp at the turnaround point (with a much smaller frequency), and in question 7 that the *acceleration* of the ball was zero at the top of the ramp. The responses from this cluster indicate a lack of logical consistency and mistaking acceleration for velocity.

While non-normative ideas were far less common overall (Table 1), our group data for the first exam in Fall 2015 (Figure 2(e)) had two prominent response clusters and shows a few connections between responses. The dark green cluster in this data represents the misconception that acceleration is $9.8\,\mathrm{m/s^2}$ at all points on the ramps; however, this time it is connected to the idea that acceleration is straight down. The other major misconception from the ramp problem, mistaking acceleration for velocity, is present in the brown cluster of the group exam data. In contrast with the blue cluster, these responses exhibit the exact same pattern as the individual exam; indicating that this misconception was strongly held. Lastly, the misconception regarding static friction was *not* present in the group data; indicating that students were able to logically resolve their ideas regarding context through collaboration.

IV. CONCLUSIONS

We used cognitive network and assessment design tools to investigate the ways in which group exams affect student thinking. Creating cognitive networks from student responses allows us to visually identify misconceptions and how they change with collaboration. One of the utilities of cognitive networks is that they may be applied to any test; regardless of whether it is given as a group exam or not. They are especially helpful in understanding how students respond to sets of contextually related questions and how students connect physics concepts within contexts.

Using MAMCR and CTT on the individual data, we found that students responded with patterns matching three common physics misconceptions. While all three misconceptions were less common in the group data, none of them

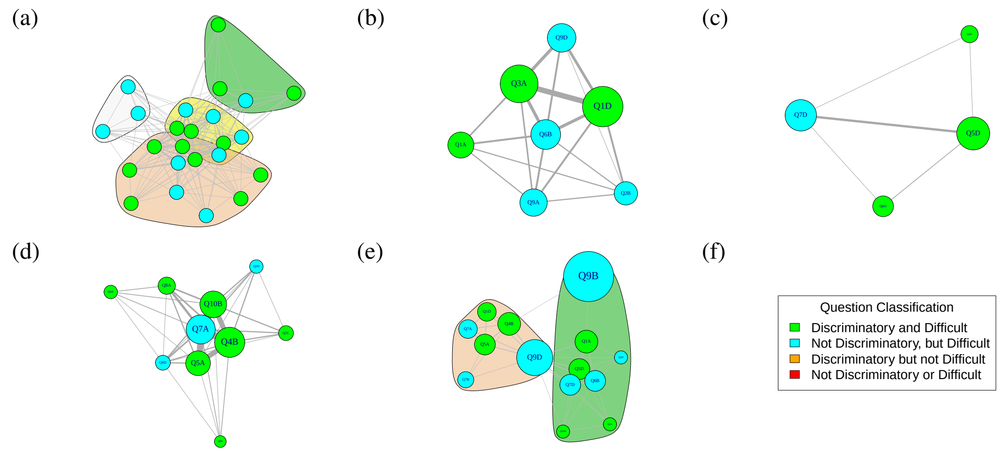


FIG. 2. (Color Online) (a) The individual test student response network for the exam under study. Each of the encircled clusters are displayed in panels b-d individually. (b) This is the yellow cluster from panel (a). It represents the misconception regarding static friction. (c) This is the dark green cluster which represents the idea that acceleration on a ramp is $9.8\,\mathrm{m/s^2}$. (d) This is the brown cluster which represents a confusion between acceleration and velocity. (e) This is the group test student response network for the exam under study. For the purposes of simplifying the view of the main network, node sizes and labels were suppressed. Also, node size is not on a shared scale across graphs. The group test student response network is also pictured(e). Node size and tie width are determined by response frequency and shared response frequency, respectively. Nodes are colored by CTT scheme (f).

were fully resolved. However, even when not resolved, incorrect responses were better logically connected to each other. Even if students answered incorrectly, their logical consistency was improved by group work. While MAMCR was previously used on the FCI, we found that it was also effective in identifying non-normative ideas and understanding how they change with group collaboration on a 'regular' test with far fewer students.

In the future, we will be using two mode and by-student MAMCR data to explore conceptual claim change during group exams. Connecting students to questions or to other students by question will allow us to analyze patterns of answer change from the individual to the group exam. We will also be working on combining multiple sets of exam data to analyze the effect group exams have on developing conceptual coherence over the long term, and exploring how the algorithm works for different test sizes.

[1] C. E. Wieman, G. W. Rieger, and C. E. Heiner, The Physics Teacher **52** (2014), 10.1119/1.4849159.
[2] B. G. W. Rieger and C. E. Heiner, Journal of College Science Teaching **43**, 41 (2014).
[3] Y. Lin and D. Brookes, AIP Conference Proceedings **1513**, 254 (2013).
[4] B. H. Gilley and B. Clarkston, Journal of College Science Teaching **43**, 83 (2014).
[5] H. Leight, C. Saunders, R. Calkins, and M. Withers, Cell Biology Education **11**, 392 (2012).
[6] S. Wolf, C. Blakeney, and H. Close, in *Physics Education Research Conference 2016*, PER Conference (Sacramento, CA, 2016) pp. 400–403.
[7] S. Wolf, T. M. Sault, and H. Close, in *Physics Education Research Conference 2017*, PER Conference (Cincinnati, OH, 2017) pp. 444–447.
[8] J. Bruun, in *Physics Education Research Conference 2016*, PER Conference Plenary Paper (Sacramento, CA, 2016) pp. 11–17.
[9] S. F. Wolf, D. P. Dougherty, and G. Kortemeyer, Phys. Rev. ST Phys. Educ. Res. **8**, 010124 (2012).
[10] S. F. Wolf, D. P. Dougherty, and G. Kortemeyer, Phys. Rev. ST Phys. Educ. Res. **8**, 020116 (2012).
[11] J. Bruun and E. Brewe, Phys. Rev. ST Phys. Educ. Res. **9**, 020109 (2013).
[12] E. Brewe, J. Bruun, and I. G. Bearden, Phys. Rev. Phys. Educ. Res. **12**, 020131 (2016).
[13] K. A. Anderson, M. Crespi, and E. C. Sayre, Phys. Rev. Phys. Educ. Res. **13**, 010121 (2017).
[14] C. Prell, *Social Network Analysis: History, Theory and Methodology* (SAGE Publications, 2011).
[15] I. Beatty, in *Physics Education Research Conference 2015*, PER Conference (College Park, MD, 2015) pp. 59–62.
[16] B. W. Frank, S. E. Kanim, and L. S. Gomez, Phys. Rev. ST Phys. Educ. Res. **4**, 020102 (2008).
[17] P. Engelhardt, in *Getting Started in PER*, Vol. 2 (2009).