SOCIO-ENVIRONMENTAL FACTORS AND THEIR IMPACT ON STUDENT COOPERATIVE ENGAGEMENT

By

JOSHUA TODD PREMO

A dissertation submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

WASHINGTON STATE UNIVERSITY
School of Biological Sciences

MAY 2019

© Copyright by JOSHUA TODD PREMO, 2019
All Rights Reserved
To the Faculty of Washington State University:

The members of the Committee appointed to examine the dissertation of JOSHUA TODD PREMO find it satisfactory and recommend that it be accepted.

_____________________________
Andy Cavagnetto, Ph.D., Chair

_____________________________
Erica Crespi, Ph.D.

_____________________________
William B. Davis, Ph.D.
ACKNOWLEDGEMENT

As with many efforts to advance one’s understanding of the world, this dissertation could not have come about in isolation. Instead, the last five years have been filled with consistent negotiation of ideas with many individuals which has resulted in the work contained herein. Foremost has been Andy Cavagnetto. A mentor whom has not just supported my research and writing, but pushed my own conceptualization of what learning is and what instruction should look like, not to impress a particular view, but to encourage a reconceptualization of what education can be… beyond what currently exists. Bill Davis has also been of great assistance over the last few years. His enthusiasm to work with me in instituting significant changes in his course have been fundamental to my ability to investigate what drives student engagement in biology classrooms and thus critical to my entire dissertation research line. Also Erica Crespi, whose open mind and consistent encouragement for my choice to complete education research, within the bounds of a largely biology focused department, deserves much appreciation. Many more individuals also deserve recognition. These include Larry Collins, Rich Lamb, Erika Offerdahl, Paul Verrell, Chad Gotch, John Hinz, Kaylan Petrie, Lisa Carloye, Ray Lee, Asaph Cousins, Kelly Puzio, and a myriad of BIOL107 TAs all of which have been important to at least one my research studies. Financial support must also be acknowledged. First and foremost has been the support of two intramural grants received from the WSU College of Veterinary Medicine that have supported this work over the last few years. Also I have received the J. Herman and Jean Kaye Swartz Graduate Fellowship, Elling Award, Charles W. and William C. McNeil Award, and a Graduate Recruitment Fellowship from the School of Biological Sciences which have contributed to my overall success in the program.
SOCIO-ENVIRONMENTAL FACTORS AND THEIR IMPACT ON STUDENT COOPERATIVE ENGAGEMENT

Abstract

by Joshua Todd Premo, Ph.D.
Washington State University
May 2019

Chair: Andy Cavagnetto

Increased attention is being paid to the role that engagement with scientific practices plays in undergraduate biology education. Particularly important, and understudied, are social aspects of scientific practice (i.e. collaboration, communication, and critique). Social aspects of scientific practice are critical to both the success of practicing scientists and those wishing to pursue a career in science, technology, engineering, and math (STEM) fields. Despite this importance, there has been limited research identifying and addressing factors that encourage students to actively engage with social aspects of scientific practice in undergraduate classrooms. Particularly, how classroom social environments differentially contribute to students’ willingness to engage with their peers. The articles herein further our understanding of how different factors in the classroom social environment relate to student willingness to cooperate with their peers during learning. First, we examined how different aspects of students’ perception of their classroom social environment relate to their dispositions towards investing in peer cooperation.
Second, we integrated a nine week intervention, designed to increase the benefit students receive from peers, into laboratory classes of a large enrollment introductory sequence biology course. Multiple data sources, including whole-classroom and individual student observations, were examined and showed that student engagement with peers increased in response to the intervention. Additional factors related to student perception of their social environment predicted variance in whole-classroom collaborative engagement beyond the intervention itself. Third, video and audio data were collected from groups completing either the intervention or a comparison activity to closely examine cooperative discourse. Differences in discourse by condition, as well as between the highest and lowest performing groups, were analyzed to provide insight into the extent to which the intervention could shift student discourse in ways aligned with greater student achievement.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENT</td>
<td>iii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>ix</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td></td>
</tr>
<tr>
<td>1. GENERAL INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2. OVERVIEW OF DISSERTATION RESEARCH LINE</td>
<td>5</td>
</tr>
<tr>
<td>3. REFERENCES</td>
<td>7</td>
</tr>
<tr>
<td>CONDITIONAL COOPERATORS: STUDENT PROSOCIAL DISPOSITIONS AND THEIR</td>
<td></td>
</tr>
<tr>
<td>PERCEPTIONS OF THE CLASSROOM ENVIRONMENT</td>
<td></td>
</tr>
<tr>
<td>1. ABSTRACT</td>
<td>10</td>
</tr>
<tr>
<td>2. INTRODUCTION</td>
<td>11</td>
</tr>
<tr>
<td>3. METHOD</td>
<td>19</td>
</tr>
<tr>
<td>4. RESULTS</td>
<td>21</td>
</tr>
<tr>
<td>5. DISCUSSION</td>
<td>27</td>
</tr>
<tr>
<td>6. REFERENCES</td>
<td>31</td>
</tr>
<tr>
<td>PROMOTING COLLABORATIVE CLASSROOMS: THE IMPACTS OF INTERDEPENDENT</td>
<td></td>
</tr>
<tr>
<td>COOPERATIVE LEARNING ON UNDERGRADUATE INTERACTION AND ACHIEVEMENT</td>
<td></td>
</tr>
<tr>
<td>1. ABSTRACT</td>
<td>37</td>
</tr>
<tr>
<td>2. INTRODUCTION</td>
<td>38</td>
</tr>
<tr>
<td>3. METHOD</td>
<td>44</td>
</tr>
</tbody>
</table>
LIST OF TABLES

1. Means and standard deviations for raw scores and correlational matrix for latent variables ......................................................................................................................................21
2. Comparison of fit for models predicting student prosocial disposition ........................................23
3. Comparison of possible models predicting student prosocial disposition ..........................................24
4. Comparison of fit for hypothesized models predicting student enforcement disposition ............25
5. Comparison of possible models predicting student enforcement disposition ..................................26
6. Description and graphical depiction of a cooperative learning session ....................................47
7. Alignment of instruments and data sources to research questions ...........................................52
8. Summary of measurement models ........................................................................................................54
9. B.E.G (Behavioral Engagement in Groups) Coding Scheme ..................................................58
10. Final regression model predicting late prosocial behavior .....................................................70
11. Final regression model predicting late discussion practices ..................................................73
12. Alignment between mechanisms of learning and discourse codes with examples of studies investigating similar aspects of group interaction (though not always relative to learning) ..109
13. Final coding scheme with embedded example ............................................................................110
14. Magnitude and alignment of discourse shifts ..............................................................................128
LIST OF FIGURES

1. Conceptual diagrams representing possible models for the impact that reciprocity, friendship, and benefit from classmates could have on student prosocial dispositions ......................................................... 16

2. Conceptual diagrams displaying representing possible models for possible factors driving student enforcement dispositions ........................................................................................................ 19

3. Cumulative structural equation model showing relationships between measured latent factors and survey items ..................................................................................................................................... 27

4. Average change in student dispositions (Willingness to Help, Reputational Concern) and perceptions of the classroom environment (Reciprocity, Friendship, Cooperative Norms, and Investment in Cooperation) from the CCEM for both comparison and intervention students ................................................................. 63

5. Comparison of overall average CCOP subscale scores by condition ........................................................................................................................................ 64

6. Comparison of longitudinal change in class level prosocial behavior between cooperative learning and comparison classrooms using the CCOP ................................................................................................................................ 65

7. Student behaviors in the comparison and cooperative learning conditions using the BEG coding scheme ........................................................................................................................................................ 67

8. Comparison of student behaviors in the intervention condition between phases 1 (normal group work) and 2 (interdependent group work) using the BEG ........................................................................ 68

9. Interaction between changes in student perception of classmate reciprocity and change in relative investment in classroom cooperation ........................................................................................................ 71

10. Mean quiz performance by condition ........................................................................................................................................................................... 118

11. Difference in the relationship between group use of justification and total content talk turns in the group by condition ...................................................................................................................................... 120

12. Predicted differences in discourse codes between interdependency (Phase 2, n = 25) and comparison (n = 28) groups ........................................................................................................................................ 121

13. Predicted differences in discourse codes between groups at different phases of the interdependency intervention (n = 25 per phase) .................................................................................................................................... 123

14. Predicted differences in discourse codes between the highest and lowest quartile of performing groups (n = 20 per each) ........................................................................................................................................ 126
DEDICATION

This work is dedicated to Megan Cartier
whose love and support over the last five years have allowed me to further myself beyond what
we could have imagined when we met in high school oh so long ago.

The, “once baby, but now toadler” Alba, whose presence has been a continuous source of
motivation over the last couple of years.

And my parents, Todd and Trina Premo, whose encouragement for self-improvement through
education has been unwavering.

I thank you all.
INTRODUCTION TO THE DISSERTATION

The goals of science education and expectations of science instruction in the United States have shifted for both K-12 and undergraduate biology education during the last decade. At the K-12 level the Next Generation Science Standards (NGSS, 2013) were developed and adopted by a number of states with the explicit goal of not only requiring students to understand specific disciplinary content, but also engage students in the “practices” of science. This emphasis on moving beyond just content understanding, as the ultimate goal of science education, is mirrored by a movement to change undergraduate biology education. This coincided with the Vision and Change in Undergraduate Biology Education: A Call to Action document. This document, and its proponents, encouraged a greater focus on the development of science competencies as a critical part of undergraduate biology education. Like with the NGSS, not all of the competencies are content based, but align instead with practices of Biologists (AAAS, 2011). The push for greater focus on science practices has opened up an opportunity for both scholars and instructors to push for greater integration of the social aspects of scientific practice into the classroom environment. This integration of the social aspects of science practice not only has the potential to increase associated student skills, but can also provide students with more epistemically authentic science experiences and greater learning.

The increased focus on practice at the K-12 level has specific expectations for what skills students should be developing, but the best practices that support student participation in this process remain unclear (Ford, 2015; Kuhn & Crowell, 2011). I would argue that a similar phenomenon has occurred in undergraduate biology education. One where the importance of competency development is understood and encouraged, but how best to support each competency remains unclear. Particularly for developing student “ability to communicate and
collaborate with other disciplines” (AAAS, 2011 pg. 15), which is the competency directly relevant to this dissertation.¹

The importance of collaboration and communication, as a coupled pair of skills, are hard to overstate. Over the last century collaboration in science has only grown in importance (Larivière, et al., 2015; Zhang, et al., 2013) and directly contributes to both the development of new understandings in science (Wang, 2016) and how successful a scientist is within their discipline (Rotolo & Messeni Petruzzelli, 2013; Sud & Thelwall, 2016). Collaboration is key to successful scientists, but also important to students seeking jobs throughout the STEM workforce (National Academies of Science, Engineering, and Medicine, 2016). Thus by supporting students’ collaborative and communication skills we can simultaneously prepare them for a diversity of jobs post-baccalaureate.

Offering students more opportunities to socially engage with their peers not only can support collaboration and communication skills, but can also offer opportunities to experience authentic social practices of science (Ford, 2008a; Hand, Cavagneto, Chen, & Park, 2016).

There are fundamental aspects to the way that science concepts arise (often through collaboration), spread through the field (via communication), and are checked for rigor (via peer review) that rely on the scientific community. The antiquated view of scientists, as individuals arriving at meaningful discoveries while working in isolation, does not reflect either the practices of current scientists nor the discipline as a whole (Ford, 2008b; Tanner, Chatman, & Allen, 2003). So making sure that student experience is not always an individual pursuit, but instead

¹ While it is certainly important to be able to work with a diversity of individuals outside of one’s discipline, I would argue that the “with other disciplines” section of this competency downplays the equally critical nature of collaboration and communication within one’s discipline and as a life competency more generally.
integrates collaboration, communication, and critique as part of the learning process, is fundamental to providing experiences that are more authentic to the social practices of science.

Both the development of social skills and more authentic exposure to social aspects of science practice are, by themselves, important outcomes. Yet when students participate in social practices of science this can also support greater student learning. There is longstanding evidence that structuring learning to be cooperative in nature (specifically via the cooperative learning model) promotes greater learning than having students work alone (for reviews see Johnson et al., 1981; Kyndt et al., 2013; Roseth, Johnson, & Johnson, 2008). Thus collaboration, in and of itself, has the potential to promote learning when structured correctly. Within the construct of collaboration there are also nuanced ties to learning through the ways that students interact with one another. Different forms of discourse are differentially related to learning (Asterhan & Schwartz, 2009; Curșeu, Chappin, & Jansen, 2018) and thus can intersect with a wide variety of mechanisms either promoting or inhibiting group learning (Nokes-Malach, Richey, & Gadgil, 2015). So having students engage in social practices of science also has the potential to promote greater learning in instances where student interaction aligns with the forms of discourse linked to learning.

While there are several potential benefits that may result from integrating social aspects of science into the classroom, they must be integrated in a way that promotes student participation. Students cannot experience social aspects of science in a classroom where active collaboration is not present. A common assumption of any activity that requires social interaction is that students will meaningfully engage with one another without additional support. Yet it has been repeatedly noted that this assumption should not be made when planning or executing instruction (Johnson & Johnson, 1999; Kreijns, Kirsch, & Jochems, 2003; Soller, 2001).
Specific approaches must be taken to encourage peer to peer interactions if we want students to experience social practices of science. Understanding the specific ways in which classroom social environments impact student willingness to engage with peers (i.e. prosocial disposition) remains opaque (Premo, Cavagnetto, & Lamb, 2018) and requires a more general examination of what drives human cooperation in order to identify factors applicable to classroom contexts.

A critical factor impacting one’s likelihood of cooperating is the benefit one receives from working with others (Kurzban, Burton-Chellew, & West, 2015). This means that increasing the likelihood of the benefit may be leveraged to promote greater student cooperative engagement. Structuring tasks to promote greater interdependency (i.e. reliance on others for one’s own success) is one method that should encourage greater cooperative benefits to be experienced by students. Interdependence can play a positive role in group learning as evidenced by its position as one core aspects of the cooperative learning model (Johnson & Johnson, 1999). Yet its role in generating prosocial classroom social environments when isolated from other aspects of the cooperative learning model is less clear.

If interdependency was able to promote more classroom environments where students are more disposed to working with their peers (i.e. prosocial), while engaging with social practices of science, then the nuances of the discourse students engage in becomes very important. This is because not all types of discourse align with either the social aspects of science practice or greater student learning (Asterhan & Schwartz, 2009). Thus the value of interdependency to science instruction is directly related to its ability to able to promote the types interactions that align with both the effective science communication and learning.
OVERVIEW OF THE DISSERTATION RESEARCH LINE

The research line presented in this dissertation includes three studies that provide insight into how we can promote greater student engagement with peers in undergraduate biology classrooms. A brief overview of these studies is as follows. The first study (Dissertation Article #1) asked the question, “To what extent do factors previously found to promote cooperation in interdisciplinary contexts predict student dispositions towards working with one another in the classroom?” The extent to which these factors were linked to students’ prosocial dispositions was examined for 845 students via structural equation modeling (Premo, Lamb, & Cavagnetto, 2018). The results of this study showed that the benefits to learning that students perceive as coming from classmates predicted their willingness to engage with peers.

Thus, an interdependency intervention was developed to promote this benefit and thus enhance student cooperation. This intervention was used in the second study (Dissertation Article #2) to address the question, “Can structuring student group work to encourage social interdependency increase student collaborative engagement?” To answer this question, the intervention was integrated into 10 laboratory sections of an introductory biology class (with an additional 10 sections for comparison) over nine weeks of a large enrollment introductory sequence biology course. Also student perceptions of their classroom social environment were collected to allow for an examination of the factors that best predicted whole-classroom collaborative engagement (Premo, Cavagnetto, & Davis, 2018). While significant increases in collaborative engagement were seen in the intervention, positive effects on student achievement were not found. This called for a more in depth examination of the nuances of student interaction within the intervention. Specifically, the types of discourse groups were engaging in and how these were related to group achievement. Thus the final study (Dissertation Article #3)
addressed the question, “Can interdependent structuring shift student discourse in ways that support learning?” This was answered through an examination of the types of discourse groups engaged in during the interdependency intervention (versus students working as they chose). Linking these differences to the discourse seen in the highest and lowest performing groups was used to assess if the intervention could shifted group discourse to promote greater learning.

It should also be noted that I, Joshua Premo, am the primary author on all of the dissertation studies featured in this document, but that each has been supported by at least one coauthor that has significantly contributed to the work. My majority contribution to published works is supported by my position as primary author on the first two dissertation articles (Premo, Lamb, & Cavagnetto, 2018; Premo, Cavagnetto, & Davis, 2018). My contributions to the unpublished third study (Dissertation Article #3) included planning, the majority of data collection, majority of data processing, sole analysis of data, and sole writer of the manuscript. All of these factors support my claim that the work represented in this dissertation is primarily my own. An additional note is that any variation in formatting between the three articles in this dissertation reflect differences in publication expectations for either where the articles were originally published (Dissertation Articles #2 and #3) or in preparation for future submission (Dissertation Article #3).
REFERENCES


A ubiquitous problem in education is getting students to cooperate effectively in groups during learning. As interdisciplinary research has found that human cooperation to be sensitive to specific contextual factors such as reciprocity, friendship, and benefit; then similar factors could be shaping student interactions in contemporary classroom environments. The present study reports on an initial investigation of this idea through examining how student perceptions of their classroom environment relate to their willingness to expend energy and resources on classmates (through helping). This was accomplished through the structural equation modeling (SEM) of survey data collected from 845 undergraduate life sciences students from 41 laboratory classrooms at a research intensive university. Results support that characteristics of the classroom social environment previously associated with benefits in other contexts also predicted (reciprocity $\beta = .576$; friendship $\beta = .156$) the educational benefit students perceive as coming from their classmates. This perceived benefit in turn was highly associated ($\beta = .605$) with student willingness to expend resources on classmates i.e. their prosocial disposition. Results also showed that both reciprocity ($\beta = .298$) and perceived benefit ($\beta = .403$) concurrently predict a student’s support for enforcing cooperation in the classroom. Potential insights into how and when student cooperation should be used are discussed.
**Introduction**

Cooperation is sensitive to the context of the situation—specifically to favor a positive cost-benefit ratio in terms of the resources expended and received (Kurzban, Burton-Chellew, & West, 2015; Nowak, 2006; Rand & Nowak, 2013; Trivers, 1971). As such, individual cooperation aligns with the expectation that, in order to maximize personal resources, individuals need to be able to selectively contribute resources to others under circumstances where a return to them is most likely. This level of selectivity is so intrinsic that even human babies as young as six months of age show preference for those persons displaying social behavior most likely to be aligned with resource return (Hamlin et al., 2007; Hamlin, 2014). By five years old, this sensitivity to the behavior of others extends to more complex decision making, with children increasing their investment in others in conditions likely to maximize future return (Engelmann et al., 2012; Engelmann et al., 2013). This sensitivity to future cooperative opportunity highlights that *even children treat others as conditional cooperators*, and for good reason. Adult studies have shown that even in complex scenarios, adults dynamically change their choice as to who they cooperate with based on an individual’s past prosocial behavior (Rand et al., 2011) and will break cooperative relationships with another person as a result of non-cooperation (Wang et al., 2012). In other words, it is important to be sensitive to how your action are perceived by potential cooperators because, if your behavior does not reflect a likelihood that their resources will be ultimately returned, you lose potential future benefit.

Educational researchers have supported these findings in noting that individuals do not interact effectively or display cooperative behaviors as a result of simply being grouped together (Johnson & Johnson, 1999; Kreijns, Kirscher, & Jochems, 2003; Soller, 2001). Indeed, several
reviews of learning with peers have highlighted the vital role of specific group contexts are in
driving learning outcomes in group learning (Kuhn, 2015; Marion & Thorley, 2016; Nokes-
Malach, Richey, & Gadgil, 2015). Cooperative learning, one of the major approaches to student
learning in groups, makes use of this conceptuality by providing specific instructional guidelines
(like structured interdependency) designed to promote greater cooperative behaviors among
students (Johnson & Johnson, 2005; Johnson, Johnson, & Smith, 2014). While both research
investigating human cooperation and student cooperation in the classroom recognize the
conditionality of cooperation, the extent to which student perceptions of their environment
impact their prosocial disposition (i.e. disposition of student towards working with others)
towards classmates remains unclear.

The current work begins examining the extent by which an optimal decision lens (where
student perceptions are aligned to increase personal benefit in the classroom) can explain a
student’s prosocial disposition and willingness to encourage cooperation among classmates.
Specifically, if a student’s perception of aspects (reciprocity and friendship) in the classroom
social environment as well as perception of benefit directly can explain student willingness to
spend time and resources helping classmates.

The Value of Prosociality and Cooperation in the Classroom

A number of studies have shown that prosociality is related to success –both
academically and socio-emotionally. For example, student prosocial behavior was found to be
highly correlated to academic achievement at the third grade (Caprara et. al., 2000) and sixth and
eighth grade as well (Wentzel et al., 2004). Prosocial student behavior has also been shown to
predict later academic achievement (Caprara et. al., 2000) and when promoted through a school
intervention programs has resulted in greater academic success than control groups (Caprara,
In addition to academic achievement, student prosociality has been shown to provide benefit to students’ socioemotional health including self-esteem (Zuffiano et al., 2014), increased student helping behaviors, and decreasing student verbal aggression (Caprara et al., 2014).

Cooperative learning research leverages student social interactions in small groups to promote academic gain. Over three decades of examining the impacts of small group cooperative learning have found that it generates significant gains in both student achievement and positive attitudes at all academic levels. In Johnson et al. (1981) in a meta-analysis of 122 North American studies found that structuring tasks in a cooperative nature in the classroom led to higher student achievement in comparison to competitive classroom environments. In 2008, Roseth et al. conducted a meta-analysis of 148 international studies of cooperative learning and also found that cooperative goals yielded positive effects sizes over both alternative (competitive $g = .46$ and individualistic $g = .55$) classroom social structuring. A final meta-analysis of 65 post 1995 studies of cooperative learning by Kyndt and colleagues (2013) further confirmed these previous findings and also reported that cooperative learning increased student achievement ($g = .54$). Both educational research into prosociality and cooperative learning show significant benefits to students in the classroom. By investigating student proclivities towards these types of social relationships, the benefits from them could be enhanced through encouraging an increased baseline of student prosocial behavior in the classroom.

**Contextual Sources of Prosociality and Enforcement of Cooperation**

**Friendships**

In order for friendships to align with an optimality lens there would have to be substantial benefits coming from friendships that outweigh the proximal costs, in time and energy, that are
associated with cultivation and maintenance of these relationships (Brent et al., 2014; Hruschka, et al., 2015). There is substantial evidence that such benefits exist including more reliable cooperation and social support, (Kurzban, et al., 2015; Massen et al., 2010) increased health, (Brown et. al., 2003; Brown et. al., 2005) and healthier children (Dunkel-Schetter et. al., 1996; Feldman et. al., 2000) which together could have dramatic positive effects on the lifetime success for individuals with long-term friendships. Even proximal educational benefits have been shown to exist with students’ friendships being associated with academic success (Wentzel, et al., 2004).

Also friendships have been shown to impact on one’s willingness to expend resources on others. Increased cooperative behavior has been shown between friends in comparison to between strangers (Majolo et al., 2006) and friendship quality has been shown to be a primary determinant of an individual’s overall social success within a group (Berndt, 2002), suggesting larger social effects arising from friendship presence. Those with friendships have also been shown to display more prosocial behavior towards others in general. This can be seen in the positive relationship between both the presence (McGuire & Weisz, 1982) and quality (Rabaglietti, et al., 2013; Sebanc, 2003) of friendships and an individual’s overall prosocial behavior in the classroom. This work suggests that classroom friends may be associated with student perception of benefit and prosocial disposition.

**Reciprocity**

Reciprocity was originally conceived by Trivers (1971) to propose a resource based reason for why individuals perform resource costly actions which do not provide immediate benefit in return. Reciprocity argues that immediate expenditures can have ultimate benefit to the actor because, although they are incurring a temporary cost for another, this cost can be
eventually outweighed by a later return from the previously helped individual. Since its original conception, the idea of reciprocity has received mathematical support for its potential role in the evolution of cooperation (Axelrod & Hamilton, 1981; Nowak, 2006). In addition there is evidence supporting the functional presence of reciprocity in human populations (Clutton-Brock, 2009; Kurzban, Burton-Chellew, & West, 2015). In quantitative field studies, reciprocity has been shown to explain cooperative food exchanges (Jaeggi & Gurven, 2013; Patton, 2005) with reciprocal cooperation being a primary predictor of meal sharing among the Ye’kwana (Hames & McCabe, 2007) and food transfers on the Ache reservation (Allen-Arave et al., 2008), Another example of these concepts in action are food exchanges in Northwest Namibia (Schnegg, 2015). Reciprocity has also been found to be a functional force when individuals interact in laboratory setting (Falk & Fischbacher, 2006; McCabe, et al., 2003) with even children as young as five showing sensitivity to reciprocal concerns (Kato-Shimizu et al., 2013). Given this substantial resource connecting reciprocity to both benefit and cooperation, student perception of classmate reciprocity would be expected to align with perception of benefit in the classroom and through this determine the student’s prosocial disposition.

**Models Predicting Student Prosocial Disposition**

Both friendship and reciprocity have empirical support connecting them to benefit and this benefit is hypothesized to drive prosocial dispositions in the classroom. The model shown in Figure 1, Diagram A illustrates that benefit is mediating the interactions between environmental factors and prosocial dispositions, making benefit the proximate gatekeeper that must be increased to affect changes in student prosocial dispositions.
An alternative view of this model may be that friendship, reciprocity, and perception of benefit can independently contribute to prosocial disposition (Figure 1, Diagram B). If supported, this alternative model would suggest that student prosociality could be encouraged by attempting to promote either classroom environmental characteristics or student perception of benefit. In addition to these models, friendship may be an important contributor to a student’s perception of classmate reciprocity. As such, this will be tested through specifying such an interaction in the models. Figure 1 C is an example of Diagram A modified to reflect specific interactions.

Each of the models were tested against one another to answer the research question, Does potential benefit from classmates mediate the relationship between social environmental conditions (reciprocity and friendship) and prosocial disposition? It was hypothesized that students will show a dispositional bias towards strategies which are most likely to maximize one’s resources. In other words, individual students will be disposed towards prosocial behavior when benefits are perceived, and be against prosocial behavior when benefits are not perceived.

![Conceptual diagrams](image)

**Figure 1.** Conceptual diagrams representing possible models for the impact that reciprocity, friendship, and benefit from classmates could have on student prosocial dispositions. A is a benefit mediated model, B predicts independent contribution to prosocial behavior, and C is a benefit mediated model with friendship impacting reciprocity.
**Models Predicting Student Disposition Towards Enforcing Class Cooperation**

Both reciprocity and friendships in the classroom could select for higher prosocial behavior in students, but their effectiveness in part, depends on individual partner choice. Partner choice allows individuals to avoid those students that do not reciprocate resources or provide immediate benefit, and has been suggested to be an important component of cooperation (Fu et al, 2008; Noe & Voelkl, 2013). Yet without this choice, enforcement of cooperation may be necessary.

Humans appear to be particularly disposed to enforce cooperation under certain conditions. For example, a human predisposition to enforce cooperation as a result of unequal behavior has been found to be common among diverse human populations from around the world (Henrich et al., 2006). Even children as young as six punish those who act unfairly (McAuliffe et al., 2015). Specific examples in traditional societies have shown enforcement to reinforce the social norms of cooperation in Ju/'hoansi Bushmen society (Wiessner, 2005) and even enforces costly cooperative warfare among the Turkana of East Africa (Mathew & Boyd, 2011). There is also significant experimental evidence supporting the impact of punishment on human cooperation. In a meta-analysis of 187 effect sizes from public good experiments (where there is a conflict between individual and collective interests and individuals have to choose one or the other) found punishment to have a substantial positive effect on cooperation ($d = .70$) (Balliet et al., 2011). This was then reaffirmed at an international scale by Balliet and Van Lange (2013) in a meta-analysis of 69 effect sizes from public good dilemmas covering 18 different societies. There is substantial variation between societies, but when taken together enforcement had a substantial positive effect on cooperation ($d = 0.77$).
Despite the ability of punishment to increase cooperation, punishment itself can be considered costly to individuals. Such a cost means that, in general, those enforcing cooperation on others would be expected to have perceived substantial potential benefits as an ultimate return from this enforcement action. Under this recognition that punishment is costly, it would be expected that students would only be willing to expend energy on enforcing cooperation (be disposed to enforce) in situations where they perceive a substantial potential benefit from classmates. A positive relationship would be expected to be present between the benefit a student perceives as coming from others and the strength of their enforcement feelings (Figure 2, Diagram D).

Yet, if the costs of punishment are divided among many group members it can become relatively less costly to any specific individual (Boyd et al., 2003, 2010; Gardner & West, 2004). In this case, students would also be expected to be more likely to have higher enforcement dispositions if they perceive the social environment to have others whom are likely to also reinforce cooperation. This focus on punishment as a group driven effect has been echoed in the experimental finding that the amount of punishment in a group is directly related to the importance of cooperation within the social group (Henrich et al., 2006). Thus if a student perceives cooperation as being common in the social group then they are more likely to enforce those not following suit (Figure 2 diagram E). Finally, students might perceive cues independently from both the potential benefit they are perceiving and group cooperative characteristics (Figure 2, Model F).

These models were tested against one another to answer the research question, Do social group characteristics (reciprocity and friendship) or potential benefit predict enforcement dispositions towards non-cooperators? Additionally, it is hypothesized that students will show a
cognitive bias towards strategies which are most likely to maximize one’s resources (be disposed towards enforcing cooperation only when benefits are perceived, and be against enforcing cooperation when benefits are not perceived) over alternative strategies (be disposed towards enforcing cooperation when benefits are not perceived, and be against enforcing cooperation when benefits are perceived).

**Methods**

**Population and Instrument**

Data was collected from 867 undergraduate students enrolled in one of 41 introductory life science course laboratory classes during the spring or fall of 2015. This was done using the Cooperative Classroom Environments Measure (CCEM) which is a 20 item classroom measure designed to assess multiple factors related to student inclinations toward prosociality and thus cooperation with classmates (Premo, Cavagnetto, & Lamb, 2018). Details as to the subscales measures and items can be seen in supplementary Table 1. Participants were given the measure 11 weeks into the semester and asked to answer items based on their three-hour per week
laboratory classroom. Of the 867 participants, twenty-two did not complete a majority of the items needed for analysis and their data was removed prior to analysis. The participants that were removed from the data set did not substantially differ from the remaining sample. This produced a final participant sample size of 845 participants. The resulting data was analyzed using Mplus Version 7.11 for structural equation modeling purposes and PASW Statistics Student Version 18.0 description statistics and chi squared analysis.

Model Evaluation

Structural equation modeling was performed on the collected data in Mplus Version 7.11 using maximum likelihood estimation. Use of the raw chi squared value is impacted by sample size (Kline, 2011) so the ratio of $x^2$ to degrees or freedom ($df$) was also examined. There is a variety of views as to what $x^2/df$ ratio is considered to be acceptable for model fit. Generally a ratio between 2 and 5 is agreed on as acceptable with lower numbers indicating better fit (Kline, 2011). For model comparison purposes the Akaike Information Criterion (AIC) was used but only when roughly comparable models existed. Typically, lower values when comparing models is an indication of more optimal fit (Burnham & Anderson, 1998). Three addition fit indices, RMSEA, CFI, and SRMR will also be reported. The root mean square error of approximation (RMSEA) was chosen as an absolute fit index with lower values indicating better fit of the proposed model to the data. For this work a strict value of < .06 is being considered as good fit for the model (Hu & Bentler, 1999). The comparative fit index (CFI) compares the specified model to the null model and is relatively insensitive to sample size (Byrne, 1998; Hooper et al., 2008). Currently a CFI of .95 indicates good model fit (Hu & Bentler, 1999). The last fit index to be reported is the standardized root mean square residual (SRMR). Values for the SRMR range between 0 and 1 with low values indicating better fit. Values for the SRMR below .05 indicate
good model fit (Diamantopoulos & Siguaw, 2000) but up to .08 is considered acceptable (Hu & Bentler, 1999).

Results

Relationships between Subscales and Measure Reliability

Means and standard deviations for all measured factors and correlations between all measured factors can be seen on Table 1.

<table>
<thead>
<tr>
<th>Variables</th>
<th>M (S.D.)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Value of Working with Classmates</td>
<td>3.61 (1.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Willingness to help Classmates</td>
<td>3.78 (0.90)</td>
<td>.591</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Reciprocity</td>
<td>3.50 (0.91)</td>
<td>.609</td>
<td>.597</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Benefit from Classmate Ideas</td>
<td>3.66 (0.99)</td>
<td>.864</td>
<td>.654</td>
<td>.669</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Friendship</td>
<td>3.13 (1.20)</td>
<td>.348</td>
<td>.282</td>
<td>.378</td>
<td>.326</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Enforcement of Cooperation</td>
<td>3.35 (1.06)</td>
<td>.590</td>
<td>.529</td>
<td>.692</td>
<td>.646</td>
<td>.296</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Benefit from Classmatesᵃ</td>
<td></td>
<td>.943</td>
<td>.711</td>
<td>.733</td>
<td>.969</td>
<td>.391</td>
<td>.709</td>
<td></td>
</tr>
</tbody>
</table>

Note. All reported correlations are significant (p < .01)
ᵃ This factor is a latent construct in common between 1 and 4.

The internal consistency and reliability of the CCEM Cronbach’s alpha value was .84, this value represents good internal consistency for study use (George & Mallery, 2003).

Additionally, the correlation between estimated and true factor scores used in the modeling was
at least .84 or above for each factor measured by the instrument showing a strong relationship between student response and their latent factor score. All items loaded onto the latent variables as expected (p <.001). Fit for the measurement model by itself was acceptable ($x^2 = 432.39$, $df = 155$, $x^2/df = 2.79$, RMSEA = .046, $CFI = .948$, $SRMR = .043$) indicating support for the presence of the factors measured by the survey.

Factors Impacting Student Prosocial Disposition

**Research Question 1:** Does potential benefit from classmates mediate the relationship between social environmental conditions (Reciprocity and Friendship) and prosocial disposition?

To provide evidence as to whether benefit mediates the relationship between student perceptions of the classroom social characteristics and their prosocial disposition, both benefit mediated (Table 2, Model A) and independent contribution (Table 2, Model B) models were specified *a priori* and examined through SEM. Results (shown in Table 3) illustrate that in comparison to the independent contribution model ($R^2 = .287$), the benefit mediated model ($R^2 = .341$) explained more variance in student prosocial disposition. Additionally, the benefit mediated model had greater fit to the collected data in each fit index. Based on these results the benefit mediated model was accepted and a final model was examined to understand whether or not friendship was impacting student perception of reciprocity. The fit of this Modified Benefit Mediate Model (Table 2, Model C) had statistically significantly better fit to the collected data. The increased fit indicates that the Modified Benefit Mediate Model best represents the contribution of Reciprocity, Friendship, and Benefit to student prosocial dispositions. Correlations were examined between friendship/reciprocity and willingness to help classmates when controlling for the effect of benefit. Results showed that the relationships between
friendship and willingness to help classmates was completely mediated by benefit (zero order $r = .282, p < .01$; controlling for benefit $r = .006, p = .85$) and that the majority of reciprocity was mediated by benefit (zero order $r = .579, p < .01$; controlling for benefit $r = .159, p = .01$).

Table 2
Comparison of fit for models predicting student prosocial disposition

<table>
<thead>
<tr>
<th>Fit Index</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>34130</td>
<td>34325</td>
<td>34077</td>
</tr>
<tr>
<td>$x^2$</td>
<td>294.65</td>
<td>490.33</td>
<td>238.09</td>
</tr>
<tr>
<td>df</td>
<td>99</td>
<td>99</td>
<td>97</td>
</tr>
<tr>
<td>$x^2$/df</td>
<td>2.98</td>
<td>4.95</td>
<td>2.45</td>
</tr>
<tr>
<td>RMSEA</td>
<td>.048 (.042 - .055)</td>
<td>.068 (.062 - .074)</td>
<td>.041 (.035 - .048)</td>
</tr>
<tr>
<td>CFI</td>
<td>.957</td>
<td>.913</td>
<td>.969</td>
</tr>
<tr>
<td>SRMR</td>
<td>.062</td>
<td>.128</td>
<td>.037</td>
</tr>
</tbody>
</table>
Factors Impacting Student Willingness to Encourage Cooperation in the Classroom

Research Question 2: Do social environmental factors (reciprocity and friendship) or potential benefit predict student willingness to encourage cooperation in the classroom?

To test this question two models were specified, one which predicted that perception of benefit would be most aligned with student enforcement dispositions (Table 4, Model D) and a second that social environmental factors (that are related to the amount of in-group cooperation) would be a better predictor of student enforcement dispositions (Table 4, Model E). Comparative results from these two models were mixed. The Perceived Benefit Model ($R^2 = .308$) was able to predict slightly more variance in student enforcement disposition, but the Social Environmental Factors Model ($R^2 = .293$) had better fit to the collected data despite having a statistically nonsignificant predictor (Table 5). This discrepancy, in addition to the finding that both models
had acceptable fit, resulted in the testing of a combined model without \textit{friendship} (Table 4, Model F). As can be seen in Table 5, this combined model had fit which was intermediate to the prior models but was able to explain 9.3\% more variance than the Social Environmental Factors model. To test whether or not this represented a significant increase in variance explained for the combined model, two regression models were analyzed (the first with perceived benefit only and the second with reciprocity and perceived benefit). Addition of reciprocity to the model resulted in a significant $F$ change value (124.79, $ P <.001$) indicating that the combined model explains significantly more variance than a single model could alone. This resulted in the acceptance of the combined model.

<table>
<thead>
<tr>
<th>Table 4</th>
<th>\textit{Comparison of fit for hypothesized models predicting student enforcement disposition}</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{Fit Index}</td>
<td>\textit{D}</td>
</tr>
<tr>
<td>$x^2$</td>
<td>180.03</td>
</tr>
<tr>
<td>\textit{df}</td>
<td>41</td>
</tr>
<tr>
<td>$x^2/df$</td>
<td>4.39</td>
</tr>
<tr>
<td>\textit{RMSEA}</td>
<td>.063 (.054-.073)</td>
</tr>
<tr>
<td>\textit{CFI}</td>
<td>.953</td>
</tr>
<tr>
<td>\textit{SRMR}</td>
<td>.049</td>
</tr>
</tbody>
</table>
**Cumulative Model Results**

When both of the best fitting models explaining student prosocial and enforcement dispositions were combined this resulted in a cumulative model (Figure 3). This model had an acceptable \( x^2/df \) of 2.82 and RMSEA value (.046) indicating close approximate fit to the data (Browne & Cudeck, 1993). The model had a CFI value of .945 which is very close to a value of .95 which is considered good fit. The final fit index used to assess the proposed model was the Standardized root mean squared residual (SRMR). Values less than .08 are consider good model fit (Hu & Bentler, 1999) so the SRMR value of .046 for the model presented here is well within the bounds of a good fitting model. All of these indices except for a slightly low CFI value support that the combined model can accurately represent the variance in student data. An additionally analysis performed to examine the roles of reciprocity and perceived benefit in mediating the relationship between friendship and enforcement of cooperation. After controlling for these factors there was no relationship between friendship and enforcement of cooperation.

**Table 5**

*Comparison of possible models predicting student enforcement disposition*

<table>
<thead>
<tr>
<th>Model</th>
<th>Predicting factor</th>
<th>Standardized Beta (b) [SE]</th>
<th>R square</th>
</tr>
</thead>
<tbody>
<tr>
<td>D – Perceived Benefit</td>
<td>Benefit from classmates</td>
<td>.675 [.072]</td>
<td>.308</td>
</tr>
<tr>
<td>E – Social Group Characteristics</td>
<td>Reciprocity, Friendship</td>
<td>.509 [.045], .087 [.045]*</td>
<td>.293</td>
</tr>
<tr>
<td>F – Reciprocity and Perceived Benefit*</td>
<td>Reciprocity, Benefit from classmates</td>
<td>.328 [.062], .366 [.061]</td>
<td>.386</td>
</tr>
</tbody>
</table>

*Note.* “Value of classmates” is a second order construct consisting of variance in common between “Value of working with classmates” and “Benefit from classmate ideas”.

* Indicates optimal model
* Factor not a significant predictor in model (p ≥ .01)
which cannot be accounted for by benefit or reciprocity (zero order r = .297 p < .01; controlling for benefit and reciprocity r = -.026, p = .450).

**Figure 3.** Cumulative structural equation model showing relationships between measured latent factors and survey items. Values are standardized parameter estimates with values in parenthesis displaying standard errors. Note: All pathways significant (p < .01).

**Discussion**

**Summary of Findings**

The current study was a preliminary attempt to understand how student perceptions of the classroom environment are related to their willingness to both work with others and encourage cooperation in the classroom. This was addressed through comparing the accuracy (via fit) of different statistical models to explain variance in student prosocial and enforcement dispositions. Overall results from SEM models showed substantial alignment with interdisciplinary studies on
cooperation. Both friendship and reciprocity significantly predicted student perception of benefit from their classmates, but with reciprocity having over a 3.5 times higher predictive power. This supports classmate reciprocity as a strong predictor of how much benefit student perceive as coming from classmates. Also a comparison between friendship’s ability to predict benefit ($\beta = .156$) versus reciprocity ($\beta = .368$) showed that friendship was a stronger predictor of benefit through increased reciprocity than to benefit directly.

In terms of student prosocial dispositions, perception of benefit was shown to mediate the relationship between friendship and prosocial disposition (willingness to help classmates) and it mediated most of the relationship between reciprocity and prosocial disposition. By itself, perception of benefit accounted for 34% of the variance in student prosocial dispositions in the final SEM model. In short, perception of benefit appears to be a key factor in determining the willingness of students to expend resources on their classmates. This suggests that instructional approaches which serve to highlight valuable student abilities and knowledge bases while buffering students with lower ability (ex: by seeding task vital content to them) are much more likely to generate student-student cooperation than simply matching higher and lower ability students. More generally, the importance of perceived benefit suggests that not every learning activity should or can simply be made “cooperative” by telling students to work together in its completion. Doing so may in fact erode students’ longitudinal perception of benefit by attempting to force them to expend extra energy and resources on cooperation when singular completion would suffice. Simply stated, cooperative learning may be most effective when there is explicit value for students as to why a cooperative approach can best meet their own learning needs. This is particularly important given repeated calls by national science education
organizations in the United States for a shift towards collaborative/ cooperative pursuits during classroom learning (Brewer & Smith, 2011; Labov, Reid, & Yokomamoto, 2010).

When student support for enforcement of cooperation were examined, the results were less aligned with the hypothesis that benefit would be driving resource expending dispositions. Instead it was found that perception of classroom reciprocity predicted variance ($\beta = .298$) in student enforcement of cooperation which was not accounted for by the student’s perception of benefit ($\beta = .403$). This finding suggests that both relative benefit as well as perceived longitudinal reciprocity may be able to independently shift the cost/benefit ratio towards or against expending resources on encouraging class cooperation.

**Limitations and Future Directions**

There are several limitations to the current study which need to be addressed in subsequent work before application. First, as a preliminary study, this work does not address actual student behavior but instead focuses on latent cognitive factors. As such, future work should explore the relationship between cognitive measures and student behavior in the classroom. Doing so may either validate survey measures as proxies for actual behavior or reveal disconnects between student cognitive dispositions and their observed behavior. Either result could have important implications for future studies interested in encouraging student cooperation and understanding the impact of classroom social environments on student actions. Additionally, the models found to best account for the relationships between student perceptions of the classroom environment and their dispositions could have been to be tested with more measures to examine convergent validity. Finally, this work does not address the impact of social expectations/ norms of cooperation on student dispositions so future work which does so may allow for the parsing of these effects in relation to student cooperation.
Conclusion

As instructors are increasingly encouraged to use cooperative and collaborative methods of instruction in the classroom, it will become apparent that, while educational research has shown that cooperative method of instruction are beneficial to the classroom, it has substantially less to say about how instructors can promote student-student cooperation. This is the vital link which connects instructional practice to student learning gains. Without knowledge of mechanisms for encouraging greater classroom prosociality, instructors have limited means of ensuring “cooperative” approaches to learning are actually cooperative in practice. The results suggest that perception of benefit is a key contributor to student prosocial disposition. Given this, strategies that make use of either prosocial or cooperative student interactions need to account for this in the classroom environment in order to maximize student interactions during cooperation.
References


PROMOTING COLLABORATIVE CLASSROOMS: THE IMPACTS OF INTERDEPENDENT COOPERATIVE LEARNING ON UNDERGRADUATE INTERACTIONS AND ACHIEVEMENT


Abstract

Collaboration is an important career skill and vital to student understanding of the social aspects of science, but less is known about relationships among collaborative learning strategies, classroom climate, and student learning. We sought to increase the collaborative character of introductory undergraduate laboratory classrooms by analyzing a nine week intervention in ten classrooms (n = 251) that participated in cooperative learning modules (promoting interdependence via a modified jigsaw technique). Students in an additional ten classrooms (n = 232) completed the same material in an unstructured format representative of common educational practice. Results showed that, when controlling for between-class variance, intervention students did not score higher on weekly quizzes, but science interest and prior science experience had a reduced relationship to quiz performance in intervention classrooms. Also intervention classrooms showed increased collaborative engagement at both whole-class and individual levels (24 students at three time points), but was only one of several factors found to account for late intervention classroom collaborative engagement (prosocial behavior and discussion practices). Taken together, findings suggest that integrating interdependence-based tasks may foster collaborative engagement at both small group and whole classroom levels, but by itself may not be enough to promote increased student achievement.
Keywords: cooperative learning; classroom environment; student collaboration; interdependency

Introduction

The ability to effectively collaborate is a necessity for success in science. Collaboration across science disciplines has grown substantially over the last century (Larivière, et al., 2015; Zhang, et al., 2013) and is able to predict both one’s contribution to science knowledge (Wang, 2016) and overall success in the discipline (Rotolo & Messeni Petruzzelli, 2013; Sud & Thelwall, 2016). This makes sense as problems in most scientific fields are too complex and large for one individual to complete alone (National Academies of Science, Engineering, and Medicine, 2016). Instruction, in contrast, tends to focus inordinately on the procedural and knowledge aspects of science (e.g. protocols, scientific method, and data analysis). While fundamentally important, these must be paired with social aspects of science (e.g. collaboration, peer review, and argumentation) to develop a complete grasp of scientific practice (Ford, 2008; Tanner, Chatman, & Allen, 2003). Yet when students experience these social aspects of science, they not only have the chance to take part in the rhetorical norms of science, but simultaneously take part in actions shown to promote immediate and delayed learning gains (Asterhan & Schwarz, 2007; 2009). Thus classroom environments that promote rich student collaborative engagement may bolster both learning and experience with the social norms of science.

Despite this potential, the factors promoting collaborative classroom environments, and the mechanisms by which they may foster greater undergraduate student learning, remain unclear. The specific contribution of collaboration to student outcomes are often hard to assess because collaboration is rarely the sole focus of an intervention (e.g. it is often paired with inquiry, Gormally, 2017; Jensen & Lawson, 2011; Seifert et al., 2009). The common practice of
introducing collaboration in tandem with more active learning practices (e.g. think-pair-share, problem based learning, inquiry instruction etc.) means that the impacts of collaborative work on both student learning and classroom environments are often entangled within broader interventions. Also cooperative learning models do not just include collaboration, but additional components, like positive interdependence, individual accountability, promotive interaction (i.e. students encouraging one another), social skills, and group processing (Johnson & Johnson, 1999), all of which could be differentially contributing to learning gains. This may be the case as cooperative learning typically has positive impacts (Johnson et al., 1981; Kyndt et al., 2013; Roseth et al., 2008), but more nuanced investigations of collaborative dynamics have shown mixed results (Barron, 2003; Chiu, 2008; Kuhn, 2015; Marion & Thorley, 2016).

The complex relationships between student collaboration and both student and classroom outcomes, as well as the consistent inclusion of collaboration alongside other interventions and variables, means that student collaboration must be more isolated to understand its impacts in undergraduate science classrooms. In the present study we evaluate how promoting student collaboration through cooperative learning modules (designed to only support one aspect of the cooperative learning model -interdependency) with a modified jigsaw technique (Slavin, 1980) impacts student achievement and interactions in the introductory life science laboratory classroom. We were interested in academic outcomes and the extent to which cooperative learning structures may be leveraged to generate whole-class communities that can effectively collaborate (Premo, Cavagnetto, & Nitta, 2017; Tanner, Chatman, & Allen, 2003). This study focuses on the following research questions: (1) Does cooperative learning (structured to encourage interdependency) impact student academic achievement? (2) Does integrating cooperative learning shift student dispositions towards collaboration or collaborative
Cooperative Learning and Student Collaborative Engagement

Cooperative learning methodologies have historically shown widespread success in promoting student achievement. Review efforts seeking to understand how cooperative structuring impacts learning can be traced back to Johnson et al.’s 1981, a meta-analysis of 122 North American studies examining the influence of cooperative learning on student achievement. The outcome of this analysis showed that structuring learning tasks in a cooperative nature resulted in higher student achievement in comparison to competitive tasks. This finding was affirmed by two more recent meta-analyses (Kyndt et al. 2013; Roseth et al., 2008). Roseth et al.’s meta-analysis included 148 international studies of cooperative learning and affirmed earlier findings by showing that cooperative goals yielded positive effects sizes over both alternative (competitive $g = .46$ and individualistic $g = .55$) goal structures in the classroom. These results were again supported by Kyndt and colleagues’ (2013) meta-analysis of 65 post 1995 studies of cooperative learning that found that cooperative learning increased student achievement ($g = .54$) to a similar extent as Roseth et al. (2008). Thus, there is significant evidence that organizing classrooms to align with the cooperative learning model can have positive impacts on student learning across a diversity of learning environments. It is crucial to note that the cooperative learning model extends beyond just having students work in groups. Students must experience positive interdependency with their peers, be held individually accountable through formal evaluation, encourage their peers to interact in a group setting, and be reflective about the processes their group goes through during learning. Concomitantly,
instructors must enforce specific student social skills (e.g. leadership, trust-building, conflict-management skills, etc.) for the traditional model of cooperative learning to be present (Johnson & Johnson, 1999). The simultaneous inclusion of all these factors makes the mechanism(s) behind student learning in these classrooms hard to isolate, and the specific contribution of student collaboration difficult to evaluate.

In contrast to the cooperative learning tradition, other lines of research investigating collaboration during learning suggest a more opaque and nuanced view of cooperation (Kuhn, 2015). For example, while cooperative learning is typically treated as a classroom level factor, it may not always be appropriate to assume that collaboration has a homogenous effect on all students. Some investigations into student-student learning have reported that student behaviors during group work can be dramatically different between groups (Barron, 2003; Chiu, 2008; Molenaar & Chiu, 2017; Sampson & Clark, 2011). This heterogeneity means that inconsistent learning gains are often found when students are asked to learn with peers (Asterhan & Schwartz, 2009; Barron, 2003; Canham, Wiley, & Mayer, 2012; Chiu, 2008; Sampson & Clark, 2011). There is also significant research noting a consistent negative effect (termed collaborative inhibition; Barber, Harris, & Rajaram, 2015; Weldon & Bellinger, 1997) of collaborative performance of groups over individual performance (i.e. when combining the contribution of each individual student independently). Importantly, even when collaboration has been found to suppress initial group performance, long-term benefits can still accrue from collaborative interaction (Blumen, Young, & Rajaram, 2014; for a review see Marion & Thorley, 2016). Thus the benefits that one receives from working with others in the classroom may only arise after repeated collaborative learning experiences around related topics over an extended timeline.
Critically, the process of simply putting students in groups does not guarantee rich collaboration (Johnson & Johnson, 1999; Kreijns, Kirsch, & Jochems, 2003; Soller, 2001). In fact, unstructured groups can have negative interpersonal effects including decreased motivation (i.e. social loafing, Meyer, Schermuly, & Kauffeld, 2015; Ying et al., 2014) and the need to contend with interpersonal conflict (de Jong, Creseu, & Leenders, 2014; Freeman & Greenacre, 2011). In other words, unstructured group work can result in negative pressures that accrue on top of the normal apprehension experienced by students participating in classrooms (Frisby et al., 2014; Howard & Baird, 2000). Traditional cooperative learning approaches explicitly recognize these sources of variance and attempt to reduce their impact on learning by purposefully structuring elements of student interaction (Johnson & Johnson, 1999; Johnson, Johnson, & Smith, 1998). These include enforcing individual accountability within the group, encouraging promotive interaction among members (i.e. students must encourage and support one another) and task framing that supports positive interdependency among group members.

All of these elements are critical to the cooperative learning model of group learning, but there is evidence that positive interdependency (simply termed “interdependency” in many cases) is especially critical to how humans collaborate across diverse social situations (Balliet & van Lange, 2013a; 2013b; Balliet, Wu, & De Dreu, 2014). Interdependency theory (Rusbult & van Lange, 2003; Rusbult & van Lange, 2008; Balliet, Tybur, & van Lange, 2016) argues that in order to understand social dynamics, one must take into account more than just the individual and overall social environment. In particular, one must also account for how interactions differ based on the degree by which needs and motives are aligned between individuals because an individual’s actions can impact another’s outcomes (i.e. they are interdependent) (Balliet, Tybur, & van Lange, 2016; Holmes, 2002; Rusbult & van Lange, 2003; Rusbult & van Lange, 2008).
Simply stated, individuals are more likely to invest time and resources into those around them when doing so is likely to encourage the attainment of their goals.

Not surprisingly, the relationship between interdependency and collaborative engagement has empirical support. Studies examining interdependency within the classroom have shown interdependency to increase observed idea generation (Buchs et al., 2010) and collaborative engagement with a learning task (Brewer & Klein, 2006; Buchs et al., 2004; Jensen, Johnson, & Johnson, 2002; Lew et al., 1986; Moser & Wodzicki, 2007). For example, Buchs et al. (2004) completed a study where students (n = 64) in pairs were either given different texts (to generate interdependency through dissimilar resources) to generate reliance on their partner or both students were provided the same text. During group work, students in the interdependent condition spent more time explaining (mr =397.9, mc = 316.7, Z = 3.15, p <.001), responding to their partner (mr =2.85, mc = 1.84, Z = 2.62, p <.01), asking questions (mr =2.96, mc = 1.98, Z = 2.34, p <.05), and making positive comments (mr =16.26, mc = 12.87, Z = 2.18, p <.05).

Additional studies have shown shifts in individual dispositions important for collaborative engagement including group self-efficacy (Alavi & McCormick, 2008), perceived social support (Bertucci et al., 2011; Bertucci et al., 2016; Johnson, Johnson, & Anderson, 1983), and perceived social cohesion (Hänze & Berger, 2007; Lew et al., 1986). Together these studies suggest that interdependency in the classroom may be leveraged to generate rich peer-peer collaborative engagement in undergraduate science classrooms. Research examining the observed impacts of longitudinal interdependency in undergraduate science education is rare. Further the research that does exist does not account for other social factors present in classrooms that have been previously linked to cooperation (e.g. friendship, reciprocity, and concern for one’s reputation, for reviews see Jaeggi and Gurven, 2013; Kurzban, Burton-Chellew, & West, 2015). From this
literature analysis, we hypothesize that integrating cooperative learning modules (designed to promote whole-class interdependency) will generate collaborative classroom environments that maximize the social dynamics needed for science knowledge construction.

Methods

Participants

Participants in the study included 483 undergraduate students who enrolled in and completed an introductory level biology course for science majors at a large research intensive university (all provided consent for participation). Participants all completed the study during a single semester. The participants were 40% freshman, 33% sophomores, 18% juniors, and 7% seniors, with the remaining 2% continuing education students. 33% of participants identified as male, 66% as female, and 1% did not identify with either gender. Participants were primarily Caucasian (66%), with 10% Latino/a, 13% Asian/Pacific Islander, 4% African American, and the remainder either of another race or they identified as multiracial. Approximately 8% of students had never completed an undergraduate science course prior to the current semester, 44% had completed 1-2 courses, 33% had completed 3-5 courses, 12% had completed 6-10 courses, and only 3% had completed more than 10 undergraduate science courses. The participating students reported a strong preference for pursuing a career in a science related field (90%); of the others 5% reported that they did not plan on pursing a science career and the remainder (5%) were unsure of their future career goals.

Design
We integrated question sets designed to support the theoretical underpinnings of a semester long CURE project (part of the SEA-PHAGES program, see Caruso, Sandoz, & Kelsey, 2009; Harrison et al., 2011; Jordan et al., 2014; Staub et al., 2016) that comprised the laboratory required for all students enrolled in a Freshman-level Biology course (Introductory Cell Biology and Genetics ). The question sets were introduced starting the fifth week of the semester and students completed a total of nine question sets over the course of the semester. The question sets focused on fundamental concepts in the SEA-PHAGES laboratory; these concepts were identified by the laboratory instructors and graduate teaching assistants as areas where previous students in the course struggled to assimilate knowledge. The themes included evaluating a scientific paper (week 1), phage- bacterial interactions (week 2), logic in experimental design (week 3), experimental troubleshooting (week 4), phages as biological tools (week 5), experimental replication (week 6), serial dilutions (week 7), plagiarism in scientific writing (week 8), and DNA isolation from phages (week 9) (see Appendix 1 in the supplemental materials). There were 21 laboratory sections of 24 students each. The exact same question set was given to every laboratory section each week. Pre-intervention, the lab sections were divided into a comparison pool (10 unique sections) and an intervention pool (11 unique sections). Since each three hour time block during the week had two laboratory sections scheduled in separate rooms, we were able to run simultaneous comparison and intervention laboratory sections to minimize variance due to the time of day that a lab was offered (except for one isolated section).

The method by which the students were asked to complete the question sets varied between the comparison and intervention laboratory sections. Students in the cooperative learning condition (intervention sections) had to work together to address core laboratory topics through structured communication with peers (see Table 1). The end goal of the intervention was
the generation of a whole-class understanding of each topic via structured task interdependency. Interdependency was encouraged by splitting how the students completed the questions sets into two distinct phases. During phase one, each group of four students (approximately 6 per lab section) were given a single question from the complete question set and were asked to work together in order to answer it to the best of their ability. This resulted in each table having specialized knowledge about a single question out of the larger set. During phase two, new groups with four students were formed so that they had within them at least one student with specialized knowledge of each question in the set. In these new groups the students were required to communicate their specialized knowledge and ideas from their first group, and then work with this second group to refine and expand on their answers to the entire question set. At the end of phase two, all groups were asked to communicate their answers to each of the questions in the set by writing them on the classroom whiteboard, followed by a discussion at the whole-classroom level designed to lead the students in the section to a common, shared understanding. This structure is similar to a traditional jigsaw technique (Slavin, 1980), which has been previously used in undergraduate science instruction (Colosi & Zales, 1998), but with two alterations. First, the structure differed in both the mechanism of original knowledge development (Phase 1 was a group instead of an individual learning alone, aligning with Doymus, 2008). Second, the level of integration was different (Phase 2 included whole-class rearrangement of individuals into new groups as opposed to individuals forming initial groups). We made these changes because the traditional design did not necessarily lend itself to whole-class collaboration due to design features best compatible for small groups. In contrast to this high level of structure, the comparison condition required each student to complete the question sets using any means they chose. There was no communication to students that they could not
work with other students, and there was no explicit offered incentive to do so. The total time dedicated to students completing each question set varied from 15 to 35 minutes. The phase three discussion (intervention only) took in additional 10 to 20 minutes. All data collection and study procedures were approved by Washington State University’s Institutional Review Board (IRB #15680) prior to initiation of the study.

Table 1

Description and graphical depiction of a cooperative learning session

<table>
<thead>
<tr>
<th>Phase 1</th>
<th>Phase 2</th>
<th>Phase 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students in each group generate specialized knowledge of a single question.</td>
<td>Students form new groups where each individual has specialized knowledge for discussion (strong interdependency present).</td>
<td>All groups participate in a whole-class discussion about the optimal answers to the question set.</td>
</tr>
</tbody>
</table>

Note. Numbers represent different students and letters represent the initial prompt students had to answer during Phase 1.

Instruments

The current study made use of five different sources of data. Data sources included: i) early and late semester measures of student perceptions of their classroom social environment (Cooperative Classroom Environment Measure [CCEM] see Appendix 2, Premo, Cavagnetto, & Lamb, 2017); ii) weekly observational rating of peer-peer collaborative interactions (Cooperative Classroom Observation Protocol [CCOP], see Appendix 3) completed by course teaching assistants; iii) six content quizzes targeting six of the nine module topics. The fourth data source was the science career motivation subscale of the SMQ-II (Glynn, 2011) to provide an initial measurement of student science career motivation prior to the study. Finally, both video and
audio recordings of group collaboration were collected from two laboratory sections (one intervention and one comparison) at three separate time points (early, midway, and late) in the intervention. The resulting video and audio recordings were analyzed using an observational coding scheme (see Coding Individual Behavior section below).

To assess students’ perception of their classroom’s social environment during the second and final week of class, each study participant completed the Cooperative Classroom Environment Measure (CCEM) – a self-report instrument that measures student perception of different aspects of the classroom social environment that can impact the likelihood of student cooperation (see Appendix 2 for instrument and Supplemental Table 1 for psychometric properties). We used a version updated from Premo, Cavagnetto, & Lamb (2017) that included subscales assessing reputational concern, perception of cooperative norms, and a student’s relative investment in cooperation. Prior work has shown that students are disposed towards cooperating with peers under conditions where their personal resources and energy are likely to be returned (Premo, Lamb, & Cavagnetto, 2017). We added the three aforementioned subscales because the initial measure could not address social pressure at both the individual (reputational concern) or class (perception of cooperative norms) levels, and had no direct measure of relative investment in cooperation (Premo, Lamb, & Cavagnetto, 2017). The important role of these subscales to the present study supported their inclusion. An examination of scale reliability for the CCEM showed that standardized Cronbach’s alpha values varied from .70-.87 for the pre-test and from .70-.85 on the post-test representing adequate to good internal consistency (George & Mallery, 2003). During the post test of this measure we also collected survey data on whether or not students planned on pursuing a science related career (Yes, No, or Undecided), gender, race, and the amount of undergraduate science courses they completed prior to the current semester.
Student interaction at the whole-class level in each laboratory section was assessed through the Cooperative Classroom Observation Protocol (CCOP, see Appendix 3 for the instrument and Supplemental Table 2 for psychometric properties). We developed this observational protocol to be a class-level assessment that teaching assistants could use to rate the nature of student collaborative interactions each week in their classroom. While the instrument targets cooperative interaction in general, we were cognizant that just because students display a positive disposition towards working with others (i.e. prosocial, see Premo, Cavagnetto, & Lamb, 2017; Premo, Lamb, & Cavagnetto, 2017) this does not necessarily mean that any student’s talk characteristics are linked to learning. For example, discussion of presented ideas (Barron, 2003; Sampson & Clark, 2011), elaborating on the ideas produced by other individuals (Barron, 2003; Stark et al., 2002; van Blankenstein et al., 2013; Van Boxtel, van der Linden, & Kanselaar, 2000), and both providing and enforcing evidence/justification for one’s ideas (Asterhan & Schwartz, 2009; Chiu, 2008; Sampson & Clark, 2011), have all been linked to learning during social interaction. Indicators of prosocial tendencies, like sharing, helping, and caring about peers (Caprara et al., 2000; Luengo et al., 2017) as well as a general preference towards social interaction (Premo, Cavagnetto, & Lamb, 2017), have the potential to exist independently from specific talk characteristics. Thus we divided the CCOP into two subscales—student prosocial behavior and discussion practices— to address both of these dimensions of cooperative interactions.

The CCOP was independently completed each week by both the graduate teaching assistants (TA; the individuals who taught the classroom) and undergraduate TAs for each lab section. One laboratory section in the intervention group did not complete at least one protocol for the majority of the eight class periods and therefore could not be included in the final analysis.
of class level behavior; this resulted in 10 comparison and 10 intervention classrooms in analyses using the CCOP. The average number of observations collected was 13.9 in each comparison classroom and 14.3 in each intervention classroom over the eight weeks. Internal consistency for each of the subscales of the observation were adequate for research purposes (Prosocial behavior $\alpha = .70$, Discussion practices $\alpha = .74$). See the Measurement Models Section below for more details on structural validity. It is important to note that these observational forms were completed by TAs during phase one of the intervention modules (where students were interacting in a small group to answer a single question) and not during phase 2 (structured interdependency) so the classroom context was as similar to that of the comparison condition as possible.

To assess students’ assimilation of the core knowledge and laboratory skills in SEA-PHAGES, six quizzes were given during the module. Each quiz contained questions whose answers were related to the information developed during the prior week’s collaborative learning exercise (see Appendix 4 for quizzes and Supplemental Table 3 for psychometric properties). TAs graded the quizzes on a zero to two point scale using guidelines provided by the course instructor. Prior to analyzing quiz scores, the scores were examined for predictive validity. The quizzes were structured to measure student understanding of science content knowledge. If scores were representative of this knowledge, then they should be related to student scores in the course overall (as this was a subsampled knowledge set). Thus, graders whose quiz scores that did not correlate to student’s overall course performance suggested that factors other than student ability primarily determined the quiz scores in that laboratory section. Correlational results showed that student overall course performance was able to predict quiz score for eight out of the ten graders. For the eight TA graders whose scores demonstrated predictive validity, there was
significant variance in this relationship (variance predicted ($R^2$) ranged from 13-43%). For the two raters whose scores did not indicate predictive validity there was no predictive ability and thus these quiz scores were not used for assessing the impacts of the intervention on student achievement. We also used confirmatory factor analysis to examine the extent to which all six of these quizzes were measuring a common underlying construct (i.e. student knowledge of the theory and practices of the SEA-PHAGES project). Results showed that scores from only five of the six quizzes demonstrated an underlying covariance structure (i.e. they were evaluating the same construct) and were therefore maintained for analysis (Supplemental Table 3). Table 2
provides an overview of each research question and the specific instruments and/or data sources used to examine results.

Psychometric Properties of Instruments

While scale reliability (Cronbach’s $\alpha$) is commonly used to assess scale quality in the social and behavioral sciences, it often misestimates actual reliability of multi-item scales (Brown, 2015). Therefore we went beyond examining reliability and assessed the quality of the measurement models of instruments using definitive anchors (CCEM, CCOP, and weekly

<table>
<thead>
<tr>
<th>Research Question 1:</th>
<th>Sub-Questions</th>
<th>Data Sources Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does cooperative learning (structured to encourage interdependency) impact student academic achievement?</td>
<td>Question 1.1: Did students in the cooperative learning condition score higher on weekly quizzes?</td>
<td>Quiz scores* ($n = 435$)</td>
</tr>
<tr>
<td></td>
<td>Question 1.2: Do students' science interest and prior science classroom experience differentially relate to performance between conditions?</td>
<td>Course performance (indicator of academic ability, $n = 483$)</td>
</tr>
<tr>
<td></td>
<td>- Prior science experience and science career interest (demographic items #36 and #37 from CCEM, $n = 483$)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Research Question 2:</th>
<th>Sub-Questions</th>
<th>Data Sources Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does integrating cooperative learning shift student dispositions towards collaboration or collaborative engagement?</td>
<td>Question 2.1: Are student dispositions towards collaboration altered by the intervention?</td>
<td>CCEM subscales (pre and post change, $n = 483$)*</td>
</tr>
<tr>
<td></td>
<td>Question 2.2: Do students in the cooperative learning condition (intervention) demonstrate increased collaboration at the whole-class level?</td>
<td>CCOP subscales (weekly whole-class observations, Appendix 3, $n = 20$)*</td>
</tr>
<tr>
<td></td>
<td>Question 2.3: Do students in the cooperative learning condition (intervention) demonstrate increased collaboration at the individual level?</td>
<td>BEG observation protocol (observations of individual student behavior, Table 3, $n = 72$)*</td>
</tr>
<tr>
<td></td>
<td>Question 2.4: Is collaboration highest when the cooperative learning activity has the most interdependency?</td>
<td>BEG observation protocol (observations of individual student behavior, Table 3, $n = 36$)*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Research Question 3:</th>
<th>Sub-Questions</th>
<th>Data Sources Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>What factors shape whole-class collaborative engagement (prosocial behavior and discussion practices) throughout a semester?</td>
<td>Question 3.1: What factors predicted late intervention prosocial behavior?</td>
<td>Late intervention CCOP subscale scores (weekly whole-class observations weeks 7-9)*</td>
</tr>
<tr>
<td></td>
<td>Question 3.2: What factors predicted late intervention discussion practices?</td>
<td>CCEM subscale measures (pre, pre-post change, $n = 483$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Initial science career motivation (SMQ II, Glynn, 2011, $n = 483$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Course performance (indicator of academic ability, $n = 483$) and rater level (undergraduate or graduate observer)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Prior science experience and science career interest (demographic items #36 and #37 from CCEM, $n = 483$)</td>
</tr>
</tbody>
</table>

* Indicates outcome examined. Additional data sources used as covariates:
quizzes) via confirmatory factor analysis (CFA). CFA is a form of structural equation modeling used to verify the latent dimensions of an instrument and it provides both evidence of structural validity as well as how these dimensions relate to the items seeking to measure them (Brown, 2015). Evaluation of the quality of a CFA measurement model is typically conducted by using fit indices that provide information about the amount of variance accounted for by the hypothesized measurement structure. Chi-squared values are reported with the recognition that, while a non-significant chi-squared value is ultimately desirable, in practice non-significance is rarely attained (for reasons like inflation based on sample size, Brown, 2015) We used three common fit indices (RMSEA, CFI, and SRMR) as well as chi-squared values to assess the quality of each measurement model. The root mean square error of approximation (RMSEA) was chosen as an absolute fit index with lower values indicating better fit of the proposed model to the data. An RMSEA < .06 was considered as a threshold value indicating a good fit for the model (Hu & Bentler, 1999) with values up to .08 being considered acceptable (MacCallum, Browne, & Sugawara, 1996). The comparative fit index (CFI) compares the specified model to the null model and is relatively insensitive to sample size (Byrne, 1998; Hooper et al., 2008). A CFI of .90 indicates adequate model fit (van de Schoot, Lugtig, & Hox, J., 2012) with .95 indicating good model fit (Hu & Bentler, 1999). The last fit index to be reported is the standardized root mean square residual (SRMR). SRMR values range between 0 and 1 with lower values indicating better fit. Values for the SRMR below .05 indicate good model fit (Diamantopoulos & Siguaw, 2000) and values up to .08 are considered acceptable (Hu & Bentler, 1999). All models were run using a robust maximum likelihood estimator to correct for variations in normality. All of the models were of at least adequate quality in all fit indices, and demonstrated good fit in at
least one fit index (see Table 3). This provided substantial support for their use in the current study.

<table>
<thead>
<tr>
<th>Index</th>
<th>CCEM (pre-test)</th>
<th>CCEM (Post-test)</th>
<th>CCOP (PB)</th>
<th>CCOP (DP)</th>
<th>CCOP (combined)</th>
<th>Weekly Quizzes</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x^2$</td>
<td>766.40*</td>
<td>853.04*</td>
<td>10.65</td>
<td>3.97</td>
<td>59.02*</td>
<td>17.52*</td>
</tr>
<tr>
<td>df</td>
<td>467</td>
<td>467</td>
<td>5</td>
<td>2</td>
<td>26</td>
<td>9</td>
</tr>
<tr>
<td>RMSEA</td>
<td>.04 (.036-.047)</td>
<td>.05 (.042-.051)</td>
<td>.06 (.000-.115)</td>
<td>.06 (.000-.153)</td>
<td>.07 (.044-.089)</td>
<td>.049 (.010-.084)</td>
</tr>
<tr>
<td>CFI</td>
<td>.92</td>
<td>.90</td>
<td>.97</td>
<td>.99</td>
<td>.93</td>
<td>.91</td>
</tr>
<tr>
<td>SRMR</td>
<td>.06</td>
<td>.07</td>
<td>.03</td>
<td>.02</td>
<td>.05</td>
<td>.04</td>
</tr>
</tbody>
</table>

* $p > .05$

**Table 3**

*Summary of measurement models*

**Addressing the Nested Nature of the Study**

Since all measures were collected from multiple classrooms, this generated a nested data structure that could have important implications for study results. Multilevel modeling approaches can be used to address nested data, but there are several criteria that collected data must meet in order for a multilevel modeling (MLM) approach to be appropriate. These criteria include 1) the data must be collected at the individual level, but be nested into larger groups, 2) sample size at level 2 (classroom) must be sufficiently large, and 3) a significant amount of the variance (normally between 10-20%) for a given measure must be able to be accounted for by between group differences (Hox, 2010). Unfortunately, since much of our data did not meet these criteria a MLM approach was inappropriate. Specifically, the data was either collected at the whole-classroom level (i.e. the CCOP observations) or it was not collected using a large enough sample size (the interaction between power and sample size at multiple levels is very complex [see Hox, 2010] with simulations commonly not conducted below 30 groups/classrooms [i.e. Maas & Hox, 2005] making data from individual coding inappropriate). Also, the between-class...
variance for student reported perceptions and dispositions were relatively minor making an MLM approach minimally meaningful (e.g. the CCEM test values where the highest amount of between group variance was only 5.9%).

The weekly quiz scores was the dataset closest to being appropriate for a MLM (we have about .65 power assuming a medium effect size according to Scherbaum and Ferreter, 2009). Given that quiz data included both individual and group levels (approximately 19 groups of 24) and 18.3% of the variance was in-between classrooms, we did use a multi-level approach to address this analysis. We specifically identified between-classroom variance by using an intercept only mixed effects model with classroom as a random effect and student-level quiz performance being the outcome as suggested by Hox (2010). This allowed the partitioning of variance as either class-level or student level (see Supplementary Table 6 for the modeling process and results). Once this was established, each student level variable (their TA, course percentage grade, and science interest) was individually entered as a fixed effect into the model and retained if it was a significant predictor of quiz performance and reduced between classroom variance (also fit statistics were examined). Intervention participation was entered into the model once the majority of between-classroom variance was statistically accounted for.

**Missing Data and Data Normality**

Missing data from the Cooperative Classroom Observation Protocol (CCOP, see Appendix 3) was examined prior to analysis. Little’s MCAR determines the extent to which data is missing completely at random (Kline, 2011). Results from Little’s MCAR test results showed that the data was missing completely at random ($\chi^2(46, N = 360) = 52.42, p = .24$) indicating that there was no systematic bias in the non-completion of the CCOP. Following this result, multiple imputation was used to account for CCOP missing data. While several researchers have argued
that a small number of imputations (3-5) are enough to generate stable results in most cases, (Schafer & Olsen, 1998) more recent work has suggested up to 100 imputations are needed to prevent power falloff using the technique (Graham, Olchowski, & Gilreath, 2007). Therefore, we created 100 imputed data sets and the results presented in this study are from their pooled results. Note that the missing quiz score data was systematic and thus could not be imputed. All collected data was first examined for normality using a Shapiro-Wilk W Test. Our results revealed that the data distributions showed only minor deviations from a normal distribution making parametric analysis appropriate. All analysis was completed using JMP PRO 12.2.0 and SPSS Version 22.

**Coding of Individual Students Behavioral Engagement**

Due to the large sample of students participating in the study only a subsample (1 class per condition at 3 time points) was recorded. These recording were analyzed for an in-depth examination of student collaborative behavior using the Behavioral Engagement in Groups (BEG) coding scheme (Premo, Cavagnetto, & Nitta, 2017). Seventy-two students were audio and video recorded while they completed the question sets (36 per condition, 12 at each time point). The audio files from students completing the questions sets for three tables (four students each) for two classes (one intervention and one comparison) over three class sessions (weeks 1, 4, and 8 of the study) were transcribed. A lack of statistical difference between subsampled sections and the larger participant pool (i.e. condition) was established by examining both students’ overall course laboratory performance and their CCEM subscale measures (see Supplemental Table 1) between subsampled classrooms in comparison to all laboratory classes of the same condition. Results showed that the classrooms chosen to be video recorded did not vary significantly in students’ course performance from the larger participant pool. There was also no statistically
reliable difference for initial CCEM subscales between the video-recorded classrooms and all classrooms of the same condition except that friendship was significantly lower in the comparison video recorded sample than the comparison population as a whole ($R^2 = .03$, minor effect). There was no significant difference between the two video recorded classrooms in terms of either academic performance or CCEM pre-test measures.

The BEG coding scheme (Table 4, Premo, Cavagnetto, & Nitta, 2017) codes the behavior of every visible student every ten seconds based on their relative engagement during group work. Engagement related behaviors varied from those representing explicit engagement with the task (discussing science content, student voicing of misunderstanding, supportive interactions) and engagement at a smaller unit level (pair talk regarding the task) to behaviors that lacked evidence for robust engagement (passively looking at the board or listening to peers) and behaviors that signified student disengagement with the task (working by themselves on another task [ex. text messaging] or being socially distracted by or with another student).

Codes were assigned based on the highest level of engagement during the ten second interval. This resulted in 2,654 coded intervals of individual student behavior ($m_{\text{student}} = 110.5$). The analysis was restricted to when groups/pairs were actively working on the question sets (first to last instance of science discussion). To test for reliable application of the BEG, all intervals were independently coded by two research assistants, and then their agreed on codes were compared to that of the primary author (who also independently coded each interval). There was strong agreement between the research assistants (Cohen’s kappa = .84) and strong agreement between the agreed codes from the research assistants and the author (Cohen’s kappa = .85). This demonstrated reliable application of the coding scheme, supporting use of the data for further
analyses. All disagreements were resolved through discussions during generation of the final data set.

Table 4
B.E.G (Behavioral Engagement in Groups) Coding Scheme

<table>
<thead>
<tr>
<th>Category</th>
<th>Behavior</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Engaged</strong></td>
<td><strong>Science content discussion</strong></td>
<td>“So, the infection will be completely treated and then the bacteriophage will go away and nothing bad will happen.”</td>
</tr>
<tr>
<td></td>
<td>Includes: Volunteering an idea (non-solicited), answering a content/concept map related question by a peer, and offering ideas within a question at the whole table level.</td>
<td>“So it's like kind of like what we're doing now, we're trying to find a bacteria phage specific to our host.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“They took the extract from the yogurt and then they added some, uh, bacteriophages that were specific to listeria…”</td>
</tr>
<tr>
<td></td>
<td><strong>Voicing of misunderstanding</strong></td>
<td>“I was like, What are they saying?”</td>
</tr>
<tr>
<td></td>
<td>Includes: Student expression of difficulty in understanding the session’s science content at the whole table level.</td>
<td>“And why did they ... did we ... centru- centrifuge the waste water sample in the first step?”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“I’ve never been a bacteria so I really don’t know how the process works”</td>
</tr>
<tr>
<td></td>
<td><strong>Supportive interaction</strong></td>
<td>“I like that idea.”</td>
</tr>
<tr>
<td></td>
<td>Includes: Interaction that add nothing to the science content specifically, but shows peer support for either each other or the task (e.g. head nods or logistics statements) at the whole table level.</td>
<td>[8 nodding head up and down following the suggestion by 2] “All right perfect.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“I will move so we can form a larger group”</td>
</tr>
<tr>
<td></td>
<td><strong>Pair talk regarding the task</strong></td>
<td>[4 to 8] “It's like, one totally kills bacteria, which obviously that would hurt your wine, but the other one alters the DNA”</td>
</tr>
<tr>
<td></td>
<td>Includes: Student discussion of the task not directed to the whole table level.</td>
<td>[14 to 11] “Why doesn't… wouldn't my body start developing resistance …?”</td>
</tr>
<tr>
<td></td>
<td><strong>Recording peer ideas</strong></td>
<td>No audible component so determined by visual examination of video recording.</td>
</tr>
<tr>
<td><strong>Neutral</strong></td>
<td><strong>Passive attention</strong></td>
<td>No audible component so determined by visual examination of video recording.</td>
</tr>
<tr>
<td></td>
<td>Includes: Students facing board, speaker, or assignment paper but not contributing to task.</td>
<td></td>
</tr>
<tr>
<td><strong>Disengaged</strong></td>
<td><strong>Explicit non-attention by ones-self</strong></td>
<td>Often no audible component so determined by visual examination. [16 has laptop open and there is audible typing] (visual examination shows the student to be typing an essay)</td>
</tr>
<tr>
<td></td>
<td>Includes: Sustained performance of a task that is not related to the concept map, includes completing a non-task assignment on a laptop, paper, or texting on phone.</td>
<td>“I hurt the tendons in them. So now a cist is forming on my hand.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“What Hogwarts house do you think you'd be in?”</td>
</tr>
</tbody>
</table>

Regression Model Selection
We used multiple regression to examine the factors shaping collaborative engagement throughout the semester. We were interested in investigating the impact of multiple sets of classroom level factors (class level CCEM subscales (Appendix 2), class level change in CCEM subscales, condition, classroom population characteristics, achievement, as well as the raters level [undergraduate or graduate TA]) on the late intervention CCOP measures (Appendix 3). Regression allowed us to examine the relative impact of these factors simultaneously (Field, 2013). Recall that the CCOP consisted of two subscales: prosocial behavior and discussion practices. A separate model selection process was completed for predicting these subscales.

Following the procedure of Jordt et al. (2017), we first generated a highly inclusive initial model for predicting late prosocial behavior and another for late intervention discussion practices. The initial regression model for predicting late intervention prosocial behavior was:

\[
\text{Late intervention prosocial behavior} = \text{rater level (grad)} + \text{initial prosocial behavior} + \text{course achievement} + \text{initial science career motivation} + \text{percentage female} + \text{percentage interest in a science career} + \text{percentage with low science experience} + \text{initial reciprocity} + \text{initial friendships} + \text{initial willingness to help peers} + \text{initial reputational concern} + \text{initial perception of cooperative norms} + \text{initial relative investment in cooperation} + \text{intervention} + \text{change in reciprocity} + \text{change in friendships} + \text{change in willingness to help peers} + \text{change in reputational concern} + \text{change in perception of cooperative norms} + \text{change in relative investment in cooperation} + (\text{change in reciprocity} \times \text{change in relative investment in cooperation})
\]

Once this initial model was established, backwards stepwise regression was used to remove factors from the model until we had the best supported model based on Akaike information criterion (AIC) validation. AIC is a relative fit index. This means that AIC values
only indicate goodness of fit when compared to other potential models, and lower values indicate a more optimal model (Burnham & Anderson, 2002). We adopted the model with the lowest AIC with the more parsimonious model (that with fewer predictors) being accepted if \( \Delta \text{AIC} > 2 > |x| > 0 \). Factors were retained in the preferred model if it met this criteria even if they were not significant predictors, as the AIC indicated that these were important for explaining the data (Burnham & Anderson, 2002). Each step in this model selection process can be seen in Supplemental Table 4. This process was repeated to examine factors predicting late intervention discussion practices but without the interaction between change in reciprocity and change in relative investment in cooperation (this interaction was only hypothesized to be important for prosocial behavior). Results from this second model selection process can be reviewed in Supplemental Table 5.

Results

Research Question (1): Does cooperative learning (structured to encourage interdependency) impact student academic achievement?

Question 1.1: Did students in the cooperative learning condition score higher on weekly quizzes? In order to assess the impact of the intervention on student achievement in the laboratory classrooms each student completed six weekly quizzes designed to measure their understanding of both the processes and theoretical underpinning of their SEA-PHAGES project. Then student scores were averaged for the five quizzes that demonstrated common factor loading (see Supplemental Table 3). A significant proportion of the variance in student quiz scores was found between classrooms (18.3% of total variance) suggesting that classroom variance needed to be accounted for before we could draw conclusions about the impact of the. Our multi-level model showed that 49% of between-class variance could be accounted for by the identity of the
TA teaching in the classroom and an additional 38% more was accounted for by the difference in student academic ability between classrooms. Interest in a science career accounted for 9.3% of between class variance. In sum, we were able to account for the vast majority of between classroom variance in quiz score interventions (see Supplemental Table 6 for detailed results). With this between-classroom variance statistically controlled for (i.e. variables were retained) we then entered condition into the model. With these controls in place condition did not predict significant variance in quiz scores.

**Question 1.2: Do students’ science interest and prior science classroom experience differentially relate to performance between conditions?** Both comparison and intervention students had approximately the same number of prior completed science courses (both 3-5 on average), level of initial science career motivation (p > .50), and course performance (p > .10) at the outset of the study. A more detailed analysis of the relationships between undergraduate science experience, science interest, gender, and quiz performance was next conducted. There was no significant relationship between science experience (number of completed courses) and quiz achievement for cooperative learning students, but a significant negative relationship between the percentage of students with low science experience (2 or fewer completed courses) and average quiz score (r = -.34, p < .05, $R^2 = .11$) in comparison classrooms. There was also a positive relationship in the comparison classrooms between the percentage of students with substantial undergraduate science experience (6 or more completed courses) and average quiz score (r = .32, p < .05, $R^2 = .10$) that was not seen in the intervention classrooms. Having a higher percentage of science career oriented students in the overall laboratory environment was predictive of increased achievement in both learning environments, but it accounted for more variance within comparison condition classrooms (r = .40, p < .05, $R^2 = .15$) than intervention
classrooms \( r = .18, p < .05, R^2 = .03 \). This observation suggests that student achievement in the intervention condition is less impacted by student demographics (prior experience in STEM; active science interest) than is student achievement in the comparison classrooms. In other words, the collaborative environment may help to “level the playing field” for students with less STEM experience and/or interest.

**Research Question (2):** Does integrating cooperative learning shift student dispositions towards collaboration or collaborative engagement?

**Question 2.1:** Are student dispositions towards collaboration altered by the intervention? To examine the extent by which student dispositions towards collaboration or perceptions of the classroom were impacted by the intervention, we first added individual pre-test CCEM (Appendix 2) subscale scores into individual regression models predicting their post-test scores for the same sub-scale. This allowed us to statistically control for initial variance in student subscale scores. Next we added condition into each model predicting post-test subscale scores. Results showed that condition did not significantly predict variance in student dispositions or perception of their classroom environment for any of the CCEM subscales \( p > .20 \), see Figure 1). This suggests that student views of the classroom environment and their dispositions towards working with peers did not shift in response to the intervention.
Figure 1. Average change in student dispositions (Willingness to Help, Reputational Concern) and perceptions of the classroom environment (Reciprocity, Friendship, Cooperative Norms, and Investment in Cooperation) from the CCEM (Appendix 2) for both comparison and intervention students. Change was determined by individual student’s post-test subscale mean minus their pre-test subscale mean. Error bars represent standard error of the mean. Condition did not predict significant variance p > .20, n = 483 students.

**Question 2.2: Do students in the cooperative learning condition (intervention) demonstrate increased collaboration at the whole-class level?** Observation scores from the CCOP (Appendix 3) were examined for the first week of the intervention and no initial differences between comparison and intervention conditions were found for either prosocial behavior (m\text{comparison} = 3.68, SE = .17, m\text{inter} = 3.66, SE = .12, t = -.061, df = 30, p = .95) or discussion practices (m\text{comparison} = 3.22, SE = .16, m\text{inter} = 3.37, SE = .09, t = .75, df = 30, p = .46) when compared using unpaired t-tests. Following this, study means for prosocial behavior and discussion practices were examined between conditions. Both prosocial behavior (m\text{comparison} = 3.53, SE = .054, m\text{inter} = 3.86, SE = .047, t = 4.522, df = 358, p < .05, d = .48) and discussion practices (m\text{comparison} = 3.07, SE = .054, m\text{inter} = 3.48, SE = .056, t = 5.296, df = 358, p < .05, d = .56) were significantly higher in the intervention condition than the comparison condition (See
Prosocial behavior also had a minor, but significant interaction with time in the intervention condition ($F(1, N = 179) = 10.3, p < .05, R^2 = .05)$ but not in the comparison condition ($F(1, N = 179) = 0.30, p = .59, R^2 < .001$) (see Figure 3). There was no significant effect of time on discussion practices in either condition. This suggests that classrooms that experienced the intervention (designed to increase interdependency) displayed more prosocial behavior and discussion practices overall, as well as longitudinal growth in prosocial behavior over the course of the intervention.

*Figure 2. Comparison of overall average CCOP (Appendix 3) subscale scores by condition. Error bars represent standard error of the mean. * $p < .05$, n = 20 classrooms*
Question 2.3: Do students in the cooperative learning condition (intervention) demonstrate increased collaboration at the individual level? Recall that students in the comparison condition could work by themselves, with a partner, or as a larger group while students in the intervention condition were encouraged to work with at least 3 other students prior to the whole-class discussion. We first examined whether or not the intervention shifted student collaboration towards group-level interaction (how they were seated in the class, tables of four students) using the BEG coding scheme (Table 4). However, it is possible that the level of collaboration would be different across phases one and two because the amount of task interdependency was designed to be higher in phase 2. Thus we independently evaluated the behaviors of the comparison group members as compared to both phase one and phase two.
intervention students. Regression results showed that student presence in a comparison classroom predicted increased time not working with their table group (four members) than the intervention students during both phase 1 ($p < .0001, R^2 = .24$) and phase two ($p < .0001, R^2 = .25$).

Next, the students’ percentage of time engaged and disengaged was compared between conditions (also BEG, See Figure 4 for comparison and Table 4 for definitions of each code). Percentages were used to normalize the data since instructors varied in the time they spent completing the question sets. Analysis was performed using unpaired t-tests with Bonferroni corrections to account for inflated alpha values. The analysis showed that students in the intervention condition dedicated proportionally more time to discussing the answers and underlying concepts of the question sets with their laboratory group ($p < .0001, R^2 = .24$) as well as being socially supportive to their peers ($p < .001, R^2 = .14$) than did students in the comparison condition. The intervention students also spent marginally more time passively listening to peer ideas ($p < .05, R^2 = .24$), and marginally less time recording peer ideas ($p < .05, R^2 = .24$) than comparison students (after using a Bonferroni correction for eight comparisons that set $\alpha = .006$).
**Figure 4.** Student behaviors in the comparison and cooperative learning conditions using the BEG coding scheme. Error bars represent standard error of the mean. Percentages were used to control for between instruct variance in time dedicated to the modules. *** p < .001 ** p < .01 * p < .05, n = 72 students

**Question 2.4:** Is collaboration highest when the cooperative learning activity has the most interdependency? Based on interdependency theory, we hypothesized that greater student collaboration (BEG, Table 4) within the intervention condition should occur when interdependency is present (i.e. during phase 2 of the module; see Table 1). Results using unpaired t-tests (with a Bonferroni correction) aligned with expectations of interdependency theory and showed that intervention students in phase two displayed significantly more time engaged with their group work overall ($p < .01, R^2 = .38$) than in phase one (Figure 5). Specifically, students in phase two spent more time recording peer ideas ($p < .001, R^2 = .34$) and
marginally more time discussing science content with their table groups ($p = .03, R^2 = .08$), passively listening ($p = .04, R^2 = .09$), and voicing misunderstanding ($p = .03, R^2 = .09$) after using a Bonferroni correction for seven comparisons that set $\alpha = .007$. No difference was found in the amount of time spent socially supporting peers or being disengaged ($p > .35$). These results show that within condition, students were more collaboratively engaged during the interdependence portion of the intervention.

Figure 5. Comparison of student behaviors in the intervention condition between phases 1 (normal group work) and 2 (interdependent group work) using the BEG. *** $p < .001$ ** $p < .01$ * $p < .05$, n = 36 students
Research Question (3): What factors shape whole-class collaborative engagement (prosocial behavior and discussion practices) throughout a semester?

Question 3.1: What factors predicted late intervention prosocial behavior? Recall that the initial regression model included initial student perceptions of the classroom environment, changes in student perceptions, classroom demographics, and intervention (among other factors; see section “Regression Model Selection” above and see Supplemental Table 4 for model selection). The final model (Table 5) included initial prosocial behavior, course achievement, condition, changes in students’ perception of classroom reciprocity, and relative investment in classroom cooperation. These factors predicted 49% of all variance in late intervention prosocial behavior (CCOP subscale see Appendix 3). The significance of initial prosocial behavior as a predictor indicates that the level of a classroom’s prosocial behavior at the beginning of the intervention was an important determinant of a classroom’s late prosocial behavior (p < .01, $R^2 = .07$). Classroom course achievement also positively predicted greater late intervention prosocial behavior (p < .01) and accounted for approximately 8% of variance in late intervention prosocial behavior. Experiencing the intervention accounted for 8% of the variance in late intervention prosocial behavior (p < .01). The remaining predicted variance ($R^2 = .26$) in late intervention prosocial behavior was determined by changes in student perceptions of their classroom social environment. Increases in classroom friendships was able to predict 3% of the remaining variance. The final predictive factors were change in student perception of reciprocity (i.e. the likelihood of a future return from those you are providing immediate support for during cooperation, Trivers, 1971), relative investment in classroom cooperation (i.e. the amount of resources/energy you believe you are dedicating to cooperation in comparison to your peers), and the interaction between these two factors. The interaction between reciprocity and investment
was included in the initial model because of its tight alignment to the expectations of interdependency theory (Rusbult & van Lange, 2003). Interdependency theory argues that individuals are more likely to invest time and resources into those around them when doing so is likely to encourage the attainment of one’s own goals. Thus when students are investing heavily into classroom cooperation, and they perceive a strong return of this investment via reciprocity, it is expected that prosocial behavior would be strongly encouraged. Our results support this hypothesis by showing that the reciprocity/investment interaction accounted for a greater proportion of variance ($p < .01$, $R^2 = .17$) than any other single predictive factor. This left the remaining 6% of variance to be accounted for by the changes in investment and reciprocity themselves.

Table 5

<table>
<thead>
<tr>
<th>Factor</th>
<th>B     (SE)</th>
<th>$\beta$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-3.65(1.40)</td>
<td>3.81</td>
<td>.07</td>
</tr>
<tr>
<td>Initial prosocial behavior**</td>
<td>.57(.21)</td>
<td>.30</td>
<td>.08</td>
</tr>
<tr>
<td>Course achievement**</td>
<td>.06(.02)</td>
<td>.33</td>
<td>.08</td>
</tr>
<tr>
<td>Intervention **</td>
<td>.16(.05)</td>
<td>.28</td>
<td>.08</td>
</tr>
<tr>
<td>Change in reciprocity**</td>
<td>1.07(.38)</td>
<td>.24</td>
<td>.02</td>
</tr>
<tr>
<td>Change in friendships**</td>
<td>.57(.21)</td>
<td>.24</td>
<td>.03</td>
</tr>
<tr>
<td>Change in relative investment in cooperation</td>
<td>.16(.53)</td>
<td>.03</td>
<td>.05</td>
</tr>
<tr>
<td>Change in reciprocity x change in relative investment in cooperation</td>
<td>17.29(4.30)</td>
<td>.38</td>
<td>.17</td>
</tr>
</tbody>
</table>

$p < .1$, $^*p < .05$, $^{**}p < .01$

Adjusted $R^2 = .49$

To demonstrate the impact of the interaction between changes in reciprocity and changes in relative investment in classroom cooperation, we modeled nine hypothetical classrooms in Figure 6. In our collected data we observed that changes in class perception of reciprocity ranged from a 0.16 decrease to a 0.14 increase (out of 5) and that changes in class investment in cooperation ranged from a 0.19 decrease to a 0.22 increase (also out of 5). To model hypothetical classrooms we used a ± 0.14 change in these perceptions since this was within the
naturally occurring range in our study. Each hypothetical classroom represents one of nine potential combinations for change in perception of reciprocity (decrease, no change, increase) and relative investment in classroom cooperation (decrease, no change, increase). All hypothetical classrooms experienced the intervention, had an initial prosocial behavior score of 3.7 (study mean), a classroom average course performance of 81.88 (study mean), and did not experience a change in friendships during the semester.

Figure 6. Interaction between changes in student perception of classmate reciprocity and change in relative investment in classroom cooperation. Each bar represents a hypothetical classroom. All classrooms experienced the intervention, had an initial prosocial behavior score of 3.7 (mean of this study), and did not experience a change in friendships during the semester. Note that decrease = -.014, increase = +0.14 on corresponding CCEM subscales (Appendix 2) and was within the range of observed changes seen in study classrooms.

Question 3.2: What factors predicted late intervention discussion practices? The final regression model predicting late discussion practices accounted for less variance (37% of total) than that for prosocial behavior (see Table 6 for final model and Supplemental Table 5 for...
model selection). Unlike the model for prosocial behavior, initial discussion practices in a classroom did not predict significant variance in late intervention discussion practices. Also, unlike the situation for prosocial behavior, the level of the rater using the CCOP (Appendix 3) had a significant effect on the scores (undergraduate raters systematically rated discussion practices higher than graduate raters, p < .01, $R^2 = .02$). Similar to prosocial behavior, academic achievement of a classroom predicted significant variance in late discussion practices, but to a lesser extent overall (p < .05, $R^2 = .01$). Initial student perceptions of the classroom environment were more important to discussion practices than prosocial behavior, with initial perceptions of reciprocity and friendship as well as initial student reputational concern accounting for 11% of late discussion practices (of these, initial perception of reciprocity contributed overwhelmingly to this amount; p < .01, $R^2 = .09$). Experiencing the intervention predicted increased discussion practices (p < .01, $R^2 = .09$), while change in student perception of the classroom environment accounted for the remaining 14% of predicted variance. Of the changes in student perception of the classroom environment (reciprocity, reputational concern, investment, and willingness to help), change in student concern with their reputation was the overwhelming contributor (p < .01, $R^2 = .07$).

When aggregated, these results suggest that several common factors contributed to both late prosocial behavior and discussion practices. These included course achievement, experiencing the intervention, reciprocity, friendship, and relative investment in classroom cooperation. The influence of the intervention on both prosocial behavior ($R^2 = .08$) and discussion practices ($R^2 = .09$) was similar, but the extent to which other factors influenced these measures varied. For example, student perception of classroom reciprocity at the start of the study was a key predictor of late discussion practices ($R^2 = .09$), but changes in student
perceptions of their classroom during the study, specifically the interaction of changes in reciprocity and investment ($R^2 = .17$), was a key predictor of late prosocial behavior. Also, student concern with how their peers are viewing them (reputational concern) appeared to function differentially by predicting decreases in student engagement in discussion practices, but was not able to predict significant variance in prosocial behavior.

Table 6

<table>
<thead>
<tr>
<th>Factor</th>
<th>B (SE)</th>
<th>β</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-8.12 (2.35)</td>
<td>3.29</td>
<td></td>
</tr>
<tr>
<td>Rater level (grad)**</td>
<td>.14(.05)</td>
<td>.23</td>
<td>.02</td>
</tr>
<tr>
<td>Course achievement*</td>
<td>.05(.02)</td>
<td>.26</td>
<td>.01</td>
</tr>
<tr>
<td>Initial reciprocity**</td>
<td>3.58(.80)</td>
<td>.57</td>
<td>.09</td>
</tr>
<tr>
<td>Initial friendship presence†</td>
<td>.44(.25)</td>
<td>.19</td>
<td>.01</td>
</tr>
<tr>
<td>Initial reputational concern**</td>
<td>-1.79(.57)</td>
<td>-.46</td>
<td>.01</td>
</tr>
<tr>
<td>Intervention **</td>
<td>.17(.06)</td>
<td>.28</td>
<td>.09</td>
</tr>
<tr>
<td>Change in perceived reciprocity†</td>
<td>1.02(.61)</td>
<td>.22</td>
<td>.01</td>
</tr>
<tr>
<td>Change in willingness to help peers**</td>
<td>2.05(.56)</td>
<td>.45</td>
<td>.03</td>
</tr>
<tr>
<td>Change in reputational concern**</td>
<td>-1.77(.46)</td>
<td>-.44</td>
<td>.07</td>
</tr>
<tr>
<td>Change in relative investment in cooperation*</td>
<td>-1.67(.66)</td>
<td>-.29</td>
<td>.04</td>
</tr>
</tbody>
</table>

† $p < .1$, *$p < .05$, **$p < .01$
Adjusted $R^2 = .37$

Discussion

We integrated cooperative learning modules into undergraduate biology laboratory classrooms to examine the specific impacts of structured interdependency (Brewer & Klein, 2006; Buchs et al., 2004; Buchs et al., 2010; Jensen, Johnson, & Johnson, 2002) on collaborative engagement and achievement during undergraduate science learning. We recognized that traditional cooperative learning models include many factors beyond student cooperation (Johnson & Johnson, 1999) and tested only interdependency in the current study. Our results demonstrate that student performance was not increased by interdependency over our common practice comparison. This finding suggests that interdependency may not drive the learning gains reported from those using a cooperative learning model. In contrast, we saw substantial shifts in
both classroom collaborative engagement and individual-level engagement in classrooms experiencing the intervention, suggesting that interdependency can be used to increase peer cooperation in classrooms.

We also found that changes in student perceptions of their classroom environment were important predictors of classroom collaborative engagement in addition to the impacts of the collaborative modules. For example, interactions between student investment in peers and reciprocity from peers (i.e. a return of investment) was the strongest factor predicting late intervention classroom prosocial behavior. This result aligns with work suggesting that students are disposed to collaborate in instances where a return is most likely (Premo, Cavagnetto, & Lamb, 2017; Premo, Lamb, & Cavagnetto, 2017), but this has now been expanded to include whole-class prosocial dynamics. This finding (in addition to other regression results) aligns with interdisciplinary work highlighting the importance of perception of benefit (for reviews see Kurzban, Burton-Chellew, & West, 2015; Rand & Nowak, 2013), reciprocity (Trivers, 1971; Jaeggi and Gurven, 2013; Patton, 2005), and friendship (Kurzban, Burton-Chellew, & West, 2015; Massen et al., 2010) in determining cooperative action. Thus, student perceptions of the ratio of relative cost/return related to cooperation in the classroom may be important for instructors to consider when designing collaborative learning environments.

Our study has both specific implications for CURE classrooms as well as for understanding cooperative learning in undergraduate science education. CUREs offer undergraduate students exposure to more authentic processes of science than are typically available in traditional labs or lecture courses and have been linked to positive changes in student self-report measures ranging from science interest and self-efficacy to their sense of belonging to the science community (for a review see Corwin, Graham, & Dolan, 2015, and a more critical
review Linn et al., 2015). Less is known about CURE impacts on student communication and collaborative practice. The closest attention to this area within CUREs are studies examining changes in student self-reported communication skills such as performance during oral presentations (Jordan et al., 2014), self-confidence when discussing science (Shaffer et al., 2014), and their ability to work as a team (Shaffer et al., 2014). The current study provides some insight into the complexity of factors impacting collaborative practice in classrooms participating in a CURE. Not only were there significant shifts in collaborative engagement in response to the intervention, but also student perceptions of their classroom environment were important predictors of the positive nature of student interaction and students’ discussion practices. These results suggest that we cannot assume that just having students work with others during CUREs is enough to generate rich collaborative interactions. We suspect that this implication applies not only to CUREs, but also to other undergraduate education settings. Supportive structures should be integrated into courses to encourage more productive collaboration if we want students to have more authentic exposure to the collaborative aspects of science.

The study results also highlight the importance of using multiple measures to holistically address observable classroom phenomena (e.g. student behavior) with multivariate contributors. By measuring collaborative engagement via multiple dimensions (prosocial behavior and discussion practices) as well as multiple levels (cognitive [self-report], individual observation, and classroom level observation) we were able to better situate the bounds in which the intervention was functioning. The triangulation between different observational measures and student self-reports allowed us to determine the impact of our intervention (8-9% of variance in collaborative engagement) relative to a diversity of other factors (e.g. academic ability, science interest, changes in student perception of the classroom, etc.). If we had only relied on student-
self reports to assess the intervention, no effect would have been found. We can also imagine the inverse occurring with students reporting changes in their abilities or practice in a classroom, with a lack of independent observational support for these changes (Panadero, Brown, & Strijbos, 2016). Therefore, future research in this area, with its accompanying complex interactions, should leverage multiple sources of data (self-report and observation).

Several limitations of the current study must be noted. First, despite these multiple sources of observed difference (classroom and individual levels), students in the intervention condition did not shift significantly in any of their self-reported views of their classroom or dispositions towards cooperation. The inability of the intervention to systematically shift student perceptions suggests that greater changes in course structure may be needed to influence student perceptions. Second, while significant work was put into examining the reliability and structural validity of the observational measure used (CCOP, Appendix 3) and a significant number of class-level observations were collected (n = 282), the measure is newly developed and must be used in a wider variety of classroom environments in order to understand both its generalizability and linkages to other features found in undergraduate science classrooms. Third, the observational analysis of video and audio recordings only included two classrooms over three time points. While we provided evidence suggesting that the two classrooms were representative of the larger samples, more examinations of student interactions are needed to support any generalizations. Fourth, the current study focused on generating greater collaborative engagement among students. The study did not specifically address shifts in specific behaviors that are particularly important to scientific practice (use of evidence and logic, providing reasoned critiques, etc.). A fruitful next step in this line of research would be to address these more nuanced aspects of students talk relative to interdependent contexts. Fifth, the weekly
quizzes used in the study were not validated prior to use and, while we presented evidence of structural and predictive validity, we have not formally evaluated either their content/face validity or difficulty level. Past work seeking to examine the impacts of cognitive load on group learning have indicated that cooperative approaches have more consistent and substantial effects when individuals cannot complete a task alone (Kirshner, Paas, & Kirschner, 2009). Thus the alignment between the task and evaluations (quizzes) needs to be increased and a better accounting for their difficulty performed during future studies.

In conclusion, structuring student learning to encourage interdependence resulted in increased collaborative engagement, but did not increase achievement. Future work should continue to address how course structures shape student collaborative interaction, and how these interactions contribute to both student achievement and their understanding science practices within undergraduate science courses. If we want undergraduate students to experience more authentic collaborative environments, and thus experience some of the most beneficial social aspects of scientific practice, we cannot assume students will do so in unsupportive settings. Class design must be purposeful and pay attention to establishing classroom environments that drive student collaborative engagement and interdependency in order to enhance student experience with collaboration as a social aspect of scientific practice.
References


DISCOURSE REMIXED: USING INTERDEPENDENCY TO SHIFT STUDENT LEARNING THROUGH TALK

Abstract

Increasing student engagement with social practices of science (communication, collaboration, and critique) is a goal throughout K-16 science education. These practices also provide an opportunity for science educators to leverage student discourse to optimize learning in their classrooms. Yet how different classroom environments promote productive forms of discourse that result to greater learning remains unclear. In the current study we examined how interdependence (via a jigsaw-like technique) shifted student discourse and thus the potential for learning. 78 groups (n_{interdependency} = 50, n_{comparison} = 28) of undergraduate students completed collaborative learning activities, either through jigsaw-like structuring or unstructured common practice, before completing a content assessment on the material the subsequent week. All groups were recorded resulting in 3,628 student talk turns that were each coded for different discourse properties. Results showed that student presence in the interdependency condition predicted significantly higher assessment performance (p < .05, R^2 = .04) beyond that predicted by their course performance overall. Examinations of discourse found that both higher performing groups, and those in the interdependency condition, demonstrated significantly (p < .007,) higher rates of providing scientifically accurate ideas (R^2 = .13 and R^2 = .05) and
justification of their ideas ($R^2 = .26$ and $R^2 = .19$). The “in common” nature of these differences suggest that the interdependency intervention may promote learning through shifting idea accuracy and justification during discourse. How this finding contributes to our understanding of discussion quality as an underlying factor impacting group learning in science is discussed.

**Keywords:** interdependency; discourse; cooperative learning; student collaboration

**Introduction**

Student engagement with the social aspects of science practice (communication, collaboration, and critique) are not only key to understanding of the epistemic orientation of science (Chinn & Malhotra, 2001; Ford, 2008; Ford & Forman, 2015; Kuhn, 2010; Tanner, Chatman, & Allen, 2003), but also directly aligned with current initiatives seeking to transform how students learn science at both the K-12 (NGSS, 2013) and undergraduate (V&C, 2011) levels. This shift towards greater focus on these aspects of science highlights the importance of language to science learning (Ford, 2015; Hand, Cavagnetto, Chan & Park, 2016), but at the same time requires a greater understanding of how best to promote student engagement with peers, and thus the discourse needed, for social aspects of science to be experienced.

This increased focus on social aspects of science practice also requires closer examinations of how students engage with peers when making sense of science ideas if we want to promote learning through discourse. Particularly how engagement with different forms of discourse may contribute to differences in student learning within cooperative classroom environments. There is evidence that the quality of student-student interaction underlies the effectiveness of active (i.e. non-lecture) approaches to teaching (Asterhan & Schwartz, 2009; Curșeu, Chappin, & Jansen, 2018). While this has been recognized, we have limited insight into
the extent to which classroom social environments drive the quality of student talk and thus learning (Howe & Abedin, 2013; Isohätälä, Näykki, Järvelä, & Baker, 2018) particularly with older students (Howe & Zachariou, 2017). This is an oversight that not only limits our understanding of the mechanisms by which peers learn from one another, but critically, how best to emphasizing social aspects of science in ways that promote student learning.

One way that recently has been used to promote greater student engagement with the collaborative aspect of science has been through encouraging interdependency via jigsaw-like structures in undergraduate laboratory classrooms (Premo, Cavagnetto, & Davis, 2018). Jigsaw-like structuring of tasks is a relatively simple intervention that has the potential to shift how reliant students are on one another and thus their collaborative engagement. This shift increases interdependency, the structured reliance on peers during learning, and has been linked to positive increases in collaborative engagement during learning (Buchs et al., 2004; Buchs et al., 2010; Premo & Cavagnetto, 2018). Shifts in the quantity of engagement may also coincide with shifts in the quality of engagement (i.e. of discourse) and thus student learning.

A prior study by the authors examined an intervention supporting student-student interdependency, via a jigsaw-like technique, and found that while both individual student and whole-class behavior was shown to shift towards significantly greater prosocial engagement and discussion practices, learning gains were not found (Premo, Cavagnetto, & Davis 2018). While not seeing achievement gains from similar interventions is not unheard of (Hanze & Berger, 2007), not witnessing a learning gain despite observed increases in the quality of whole-class student engagement may highlight the dependency of learning on smaller scale contextual discourse (Howe & Zachariou, 2017; Kuhn, 2015). Thus a closer examination of the relationship
between group discourse and learning was needed to determine the modes of discourse that best promote learning in groups.

This study represents the next step into this exploration. One where the outcome measures were increased in difficulty (though the questions remained thematically identical, see comparisons in Supplemental Figures 1-4 in Appendix 2) and nuances of group discourse was the main focus of analysis. The main research question was, “Can interdependent structuring shift student discourse in ways that support learning?”. To answer this question, video and audio recordings of student interactions (i.e. talk turns) from 78 groups of students were analyzed while they completed group work in either an interdependency condition (25 groups per intervention phase, 50 total) or in the comparison condition (28 groups) where students had control over their method of completion. All student exchanges where transcribed and coded (n = 3,628) by different aspects of the discourse, at the talk turn level, before analyses were completed. This analyses sought to answer the following research questions: 1) To what extent did student participation in the interdependency condition promote learning? 2) How did interdependent structuring shift group discourse?, and 3) How does discourse vary between the highest (4th quartile) and lowest (1st quartile) performing student groups? Commonalities in discourse shifts between research questions 2 and 3, in alignment with past research into mechanisms of group learning, will then inform the overall answer to the question, “Can interdependent structuring shift student discourse in ways that support learning?”.

**Literature Review**

**Promoting Collaborative Engagement through Interdependency**

Grouping students does not guarantee effective interactions (Johnson & Johnson, 1999; Kreijns, Kirsch, & Jochems, 2003; Soller, 2001). Asking students to work in groups can
actually result in negative student experiences and learning in groups. For example, recent work has highlighted that students can simply choose to not interact or remain disengaged in the group overall (i.e. social loafing) even in higher performing groups (Chang & Brickman, 2018). In addition, a number of barriers exist, including exclusion by peers and individual anxiety, which can negatively shape how students engage in discussion (Eddy et al., 2015). The potential for negative effects and the lack of guaranteed benefits from group collaboration (Howe & Zachariou, 2017) require that an evaluation of the relative costs and potential benefits to learning should take place when deciding whether to implement collaboration in the classroom (Andrew & Rapp, 2015). In an attempt to shift this balance towards greater benefits, cooperative learning methods provide explicit structure to how students work with one another.

Cooperative learning structures student interaction by enforcing individual accountability, providing support for social skill development, requiring promotive interaction (i.e. encouragement between students), and structuring tasks as interdependent (Johnson & Johnson, 1999). Implementation, or support, of each factor is important to using the cooperative learning model and most likely contributes to the overall success, in terms of learning, of cooperative learning (Johnson et al., 1981; Kyndt et al., 2013; Roseth et al., 2008) despite the potential negative effects of cooperation. Yet it is not always practical for an instructor to integrate all structures of the cooperative learning model. For example, elements like promotive interaction requires examination of student talk to assess presence, and consistently focusing on social skill development may not be something that those instructing adult students (i.e. undergraduates) may be willing to commit to. Yet structuring tasks to promote interdependency often can be easily integrated into existing activities and curricula through the division of resources and/or task responsibility and still provide benefits to group cooperation and potentially learning.
Interdependency has been repeatedly shown to promote collaborative engagement between individuals across a variety of contexts (Balliet & Van Lange, 2013a; 2013b; Balliet, Wu, & De Dreu, 2014). This finding has led to the argument that one must understand more than the immediate social interactions between individuals, but also how aligned the goals and needs between these individuals are. This is because in cooperative situations, one’s actions can impact the likelihood of others achieving their goals which can alter cooperation behavior (Aktipis et al., 2018; Balliet, Tybur, & Van Lange, 2016; Gerpott et al., 2018; Rusbult & Van Lange, 2003).

There is evidence that students do vary in the value they place on peer interactions during learning (Chang & Brickman, 2018; Eddy et al., 2015; Premo, Lamb, & Cavagnetto, 2018) as expected by interdependency theory, but whether interdependency is driving some of this variation is unclear. Manipulations of interdependency (e.g. under jigsaw-like structuring) offer a direct way to alter the value of peers and thus test potential effects on group engagement.

Several studies that have examined interdependency have shown increased cooperative engagement (Brewer & Klein, 2006; Buchs et al., 2004; Jensen, Johnson, & Johnson, 2002; Lew et al., 1986; Moser & Wodzicki, 2007; Premo & Cavagnetto, 2018) supporting the role of interdependency in the classroom.

While interdependency theory provides a mechanism to promote collaborative student engagement, less is known about how interdependent structures shift the quality of student discourse and thus learning. Some studies have noted general shifts in discourse including more science discussion and social support of peers (Premo & Cavagnetto, 2018) as well as increased idea generation (Buchs et al., 2010), and greater whole-class engagement in discussion practices (Premo, Cavagnetto, & Davis, 2018). Yet links between these shifts and group learning were not well established and the discourse practices examined were limited in scope. Thus a more
inclusive examination of discourse would be informative to our understanding of how interdependency may shift dialog between students. Particularly if these shifts can be linked to differences in group learning.

**Learning in groups**

Research across both cognitive psychology (Blumen, Young, Rajaram, 2014; Nokes-Malach, Richey, & Gadgil, 2015) and the learning sciences (Asterhan & Schwarz, 2016; Howe, 2014; Kuhn, 2015) have repeatedly noted that we lack a robust understanding of how interactions in groups shape cognition and thus group learning. This limits our ability to optimize learning in groups. Understanding how to optimize group learning is particularly important given that research has found that some groups perform worse than expected (termed “collaborative inhibition”; Barber, Harris, & Rajaram, 2015; Weldon & Bellinger, 1997; Marion & Thornley, 2016) and that working in a group can lack an advantage to students for some outcomes like producing scientific arguments (Sampson & Clark, 2009) or mastery and transfer performance in physics (Zu, Munsell, & Rubello, 2019). In order to address this gap there has been an increased focus on the conditions by which working in groups result in learning, and a variety of mechanisms have been suggested to account for differences in learning among groups.

One method of addressing this gap has been the purposeful manipulation of different group properties and examination of differences in group learning. These have suggested a number of potential mechanisms that may drive student learning in groups (Nokes-Malach, Richey, & Gadil, 2015). Mechanisms include increased working memory resources (Kirschner et al., 2018), greater opportunities for information retrieval either through reiterative exposure to ideas (Rajaram & Pereira-Pasarin, 2010) or cueing from other members of the group (Congleton & Rajaram, 2011), and increased opportunities for error correction (Ross et al., 2004). All of
these factors may play a role in groups learning, but vary in degree based on the immediate context. It is also important to note that a common outcome measure for these studies is individual recall of item lists (Marion & Thorley, 2016). While these allow for straightforward interpretation of outcomes and serve to minimize the influence of prior knowledge, this outcome may not always translate to authentic classroom outcomes which can be much more complex in nature.

Other approaches to examining how learning occurs in groups takes a more exploratory approach. For example, by examining differences in the ways in which groups interact to see if these can account for differences in learning. Some studies focus on specific interactions and its variance among all groups to see if there is a relationship to achievement (Chiu 2008a; 2008b; Christian & Talanquer, 2012; Clark & Sampson, 2008; Haussmann, Chi, & Roy, 2004; Purzer, 2011; Sampson, Grooms, & Walker, 2011). Others use an outcome measure to define higher and lower performing groups which are then explicitly compared to see how group interactions differ (Asterhan & Schwartz, 2009; Barron, 2000; Barron, 2003; Evagorou & Osborne, 2013; Khosa & Volet, 2014; Sampson & Clark, 2011). These differences are then proposed as potential sources of learning. The diversity of these studies has resulted in a variety of potential factors that could be driving learning in groups.

One factor commonly found to be related to learning is the quality of group discourse (Howe, 2014; Howe & Abedin, 2013). A greater understanding of this relationship provides a window through which we can gain a better understanding of potential mechanisms underlying how students learning in groups. Yet not all mechanisms of group learning are best be examined through analysis of discourse. Some for example, like the presence of a collaborative cognitive load (i.e. the ability of groups to be able to process larger amounts of more complex information.
than an individual can alone) may increase the working memory resources available to groups and thus allow them to outperform individual during complex tasks (Kirschner, Paas, & Kirschner, 2009; Kirschner, Sweller, & Kirschner, 2018), but how this may be examined via discourse is unclear.

**Learning through discourse**

Other mechanisms of group learning lend themselves to examinations of discourse. Four that have some overlapping features, but distinct mechanisms, will be the focus of this study. First, increased opportunities for information retrieval can promote learning in groups (Congleton & Rajaram, 2011; Rajaram & Pereira-Pasarin, 2010). Repeated retrieval of information from memory has long been studied as a mechanism to promote retention of ideas (Cepeda et al., 2006) and discussion in groups both provides explicit opportunities for student to recall their understanding both individually and relative to others. Second, having at least one other individual to talk with increases the chance of errors being corrected (Ross et al., 2004) thus pruning of ideas towards having greater accuracy. Greater accuracy increases the likelihood of a more accurate answer being arrived at by the group which can be key to subsequent performance (as in many cases immediate group performance is critical to subsequent recall, for a notable exception see Kapur, 2008; Kapur, 2010). Also groups can have an increased accuracy in comparison to individuals working alone (Bol et al., 2012; Sniezek & Henry, 1989) which may privilege group learning over individuals in general. Third, comparing ideas between different perspectives (Andrews & Rapp, 2015; Asterhan & Schwarz, 2016) can force students to reevaluate their existing knowledge structures and thus reinforce existing knowledge structures or promote conceptual change (Chi, 2008). This reinforcement/conceptual change process can result in either more robust or changed student schema that impacts subsequent retrieval. Fourth,
the increased need for explanation/justification of ideas in group contexts, when present, can reinforce student’s knowledge structures through the explicit need to formalize and communicate these ideas in ways understandable to peers. This process has been related to greater learning (Webb, 1991). All of these potential mechanisms could be driving group learning, but what evidence is there for each?

**Opportunities for information retrieval.** Working with peers offers the potential for students to be re-exposed to concepts and ideas during discourse through cueing repeated retrieval. Increased opportunities for retrieval have the potential to improve learning (Blumen & Rajaram, 2008, Congleton & Rajaram, 2011; Meudell, Hitch, & Boyle, 1995). This process (termed cross-cueing when causes by another group member) occurs when the ideas of one individual cue recall in others and thus result in retrieval practice. Retrieval in groups appears to be qualitatively different from individual recall as it has been shown that group retrieval can result in maintaining individual recall over longer delays and demonstrates greater retrieval organization (indicated by adjusted ratio of clustering scores) in comparison to an individual performing the same retrieval by themselves (Congleton & Rajaram, 2011). Group retrieval has also been shown to coincide with a decrease in unique recall being present as shared schema becomes more formalized between members (Blumen, Young, Rajaram, 2014). In contrast to these findings, Zu, Munsell, & Rubello, (2019) found that promoting individual retrieval of information can be more effective than the retrieval that naturally happen in groups. This suggests that group retrieval may not be as optimal as structured individual retrieval in some contexts.

One proxy for examining the potential for retrieval opportunities is the amount of discussion that occurs within a group. Under this case we would expect more discussion to be
related to greater learning through increased opportunities for retrieval. This corresponds to the finding that increased time spent developing a solution (i.e. more opportunities for retrieval) in a group is related to greater student transfer ideas (ability to apply ideas out of the learned context; Canhan, Wiley, & Mayer, 2012). Yet the difference in learning from the prior study could also have been driven by differences in background knowledge as well as increased solution development due the purposeful structuring of individuals with diverse training. This interpretation is supported in the context of mathematical problem solving where the amount of total talk has been found to be unrelated to achievement (Barron, 2003). Yet in chemistry it has been found that higher performing groups can generate twice the number of unique ideas offering more diverse opportunities to cue retrieval and thus learning (Sampson & Clark, 2011). Taken together there is evidence that retrieval opportunities in groups may be related to learning, but the extent to which this varies by context and is optimal remains to be seen.

**Increased accuracy in groups.** Another potential mechanism that could promote learning in groups in increased accuracy via error correction (Ross et al., 2004). Just looking at the contribution of accuracy to learning in groups, studies have found that the number of accurate contributions (Chiu, 2008a) as well as unique accurate contributions (Chiu, 2008b) in a group are related to subsequent achievement. Also higher achieving groups have been found to be more likely to confirm correct ideas when presented to their group (Barron, 2000) supporting the differential treatment of accurate/inaccurate ideas among higher achieving groups. The process of error correction is more complex than the relative presence/absence of accurate ideas. This is because error correction can take many different forms within group discourse. At a basic level, error correction can be examined simply by examining how a group responds to ideas in the group with accurate ideas being accepted and inaccurate ideas resulting in disagreement by
peers. The acceptance of ideas by higher achieving groups has been found to differ by context with higher achieving groups displaying greater acceptance in 6th grade math problem solving (Barron, 2003), but less acceptance of ideas in undergraduate science learning (Sampson & Clark, 2011). Yet in both cases there was more discussion of these ideas in higher performing groups suggesting that the discussion process may be a more important contribution to learning than just acceptance by itself (Sampson, Grooms, & Walker, 2011).

Idea comparison. Comparing ideas between different perspectives (Andrews & Rapp, 2015; Asterhan & Schwarz, 2016) has the potential to force students to reevaluate their existing knowledge structures and thus shift learning. We would expect idea comparison to be particularly important when inaccurate ideas are presented to a group. Higher performing groups would then be expected to present a challenge or disagreement with these ideas. Yet this is not always the case and can vary based on how the disagreement is phrased. Disagreement has been shown to positively predict the likelihood of a new idea occurring if phrased politely, but decreases the generation of a new ideas is phrased rudely (Chiu, 2008b). Outside of its rude/polite phrasing, disagreement can result in idea comparison where students attempt to resolve presented discrepancies through critique. While this persuasive process has not always been linked to achievement (Barron, 2003; Purzer, 2011), it has been in a variety of contexts. For example, the presence of critical aspects of idea comparison have been noted as a defining feature of high achieving groups at the middle school (Evagorou & Osborne, 2013), high school (Sampson & Clark, 2011; Sampson, Grooms, & Walker, 2011) and undergraduate (Asterhan & Schwartz, 2009) levels.

Comparing one’s ideas to another involves not just disagreement, but additionally pushes individuals to explain and justify their ideas in the process of resolving discrepancies.
Unsurprisingly then, is the finding that individuals are more likely to provide justification (via evidence) in the presence of critical discourse (Clark & Sampson, 2008) and that comparing one’s ideas has been found alongside justification as a practice of higher performing groups (Sampson & Clark, 2011). The comparison and explanation process can also result in idea negotiate that promotes more effective integration of ideas, counter ideas, and perspectives (Kuhn & Crowell, 2011). Yet how much critical feedback itself drives these outcomes versus the explanatory process remains unclear.

**Justification of ideas through explanation.** Justification of ideas through explanation can reinforce student’s knowledge structures through the explicit need to formalize and communicate ideas to their group. Yet the extent to which idea comparison itself is driving learning during critical discourse, versus the co-occurrent increased need for verbal justification and explanation of ideas is difficult to differentiate. For example, Asterhan & Schwartz (2009) in their examination of discourse factors that differed between learning (pre-post gain) and non-learning pairs, found that total amount of dialectic argumentation (included oppositions, challenges actual rebuttals and concessions) significantly differed when combined. Yet analysis of individual discourse move differences (uncombined) revealed that, of these, only actual rebuttals (which they defined as a response to a challenge to strengthen their explanation) was the predominant discourse move which varied between groups. Given these results, it may be reasonable to conclude that need for explanation may have been a prime driving factor that differentiated between learning and non-learning pairs and not the critique process itself.

The relationship between explanation and learning does align with a number of older studies attempting to identify which behaviors are related to learning in groups. Webb (1991) noted, in a review of fifteen studies looking at the relationship between group behaviors and
achievement, that in 11/15 cases giving elaborated explanations in groups were positively related to achievement ($r = .22-.47$) when controlling for academic ability. This suggests that groups providing more detailed explanations for ideas will perform better than those that do not. It should also be noted that none of these studies showed any relationship of answer giving (without explanation) for those providing an answer and 5/8 showed a negative relationship ($r = -.26$ to -.50) between receiving an answer and achievement. Together these results suggest that explanation during discourse is particular important for achievement in dialogic settings and that this relationship is not resulting from simply the answer that is part of the explanation but the cognitive process of generating the explanation itself.

There are several more recent studies that also support the importance of explanation to learning in groups and provide insight into how learning benefits may arise. For example, self-directed explanation in groups has been found to have greater learning gains for the speaker (71%) than the listener (29%) (Hausmann, Chi, & Roy, 2004) further supporting explanation as a factor promoting learning particular for the person executing the discourse (i.e. not benefiting the group as a whole as much as those performing it). Even if performing an explanation preferentially benefits the speaker, explanation may also trigger other mechanisms that can promote learning for the group. For example, providing an explanation, via justification, has been found to be a predictor of the subsequent generation of accurate ideas in a group. Using statistical discourse analysis, Chiu (2008a) was able to show that justification was one of the strongest predictors of subsequent accurate contributions coming about in group. The amount of accurate contributions were then able to predict higher final solution scores (a measure of achievement). The ability of justification to predict accurate contributions also functioned differentially between higher (+68% chance) and lower performing (+29%) groups suggesting
that explanations can function differently, within the flows of group discourse, among groups and may contribute to differential group learning.

In summary, there is evidence supporting that opportunities for informational retrieval, accuracy of ideas, idea comparison, and the explanation/justification of ideas may all be contributing to effective discourse in at least some groups. Whether or not one or more of these are related to learning in undergraduate biology groups and the extent to which interdependency may be able to promote these discourse factors is the subject of the current study.

**Method**

The current study examined undergraduate student group work in the laboratory classrooms of an introductory biology course. Specifically at two time points (weeks two and five) of a larger nine week intervention seeking to increase students prosociality and cooperative engagement in the course (for a more in depth treatment of the entire intervention see Premo, Cavagnetto, & Davis, 2018). Each week students had to answer three questions focused on different themes of their course-based undergraduate research experience (part of the SEA-Phages program, see Caruso, Sandoz, & Kelsey, 2009). These were largely identical to those seen in Premo, Cavagnetto, and Davis (2018) with only minor wording variation (See Supplemental Figures 5 and 6 in Appendix 2 for the question sets used).

The questions answered were the same for all groups, but the method by which students completed the questions varied by condition. Students in the interdependency condition had a pre-specified format by which they completed the questions each week. This was a jigsaw-like format where individual students developed specialized knowledge of a single question in a group (Phase 1) before coming together with a new group to collaboratively address all questions (Phase 2). This structure differed a bit from other jigsaw approaches by having students work
collaboratively instead of individually in the initial phase. This variation was chosen to provide students more opportunities to discuss with their peers and thus increase the quality of both student discussion and understanding prior to entering their Phase 2 groups- which included at least one students with specialized knowledge of each question from Phase 1 (see Premo, Cavagnetto, & Davis, 2018 for more details on this structure). The comparison groups did not include explicit structuring. Though students were seated in tables that included four students in both conditions, comparison students could complete the questions in any format they chose (e.g. by themselves, with their group, or with just a partner). No specific format was encouraged by the instructors. Instructors were asked to encourage students to rely on their peers for answers in both conditions, but not the instructors themselves as the focus was on group discourse. Instructors were largely successful in this regard with only 36 student talk turns directly resulting from student interactions with an instructor (out of 3,628 total coded student talk turns). After completing the questions, all students in both conditions took part in a class level discussion about the answers to the questions (Phase 3).

All students in the course were asked for consent to be video/audio recorded and eight classrooms with at least 75% consent were chosen for video/audio recording. Due to two simultaneously running sections of the laboratory class, only one of these could be audio/video recorded during each time period due to equipment and personnel limitations. The one chosen was simply based on which had higher student consent. Condition was assigned at the class section level (5 interdependency, 3 comparison) and if an instructor taught more than one recorded class, these were all of the same condition to reduce blending between intervention and comparison formats. There were seven unique instructors for the laboratory classes with two instructors teaching two of the classes where data was collected and the remaining only teaching
one class each. The classrooms included tables along both the left and right sides of the room that could seat four students each. When recording was to occur in the interdependency condition, instructors asked that all students that had consented sit in the right tables of the room and these were recorded. This arrangement allowed for an element of random sampling to occur as only half of the students in a class were recorded, but students were asked to stay on the sampled side of the room in the switching between Phase 1 to Phase 2 to prevent the recording of non-consenting students. Sampling in the comparison condition worked in a similar way by starting with the right three tables in the room. Comparison students did not participate in the rearrangement of groups, and had reduced potential of non-consenting students being recorded. This allowed additional tables to be recorded if fully consenting groups were present on the left side of the room. This increased the total number of groups from the comparison condition.

Redundant recording devices were used to capture student talk due to the high volume and overlap of conversations as whole classes were actively engaged in discussion. Each group was recorded by their own video camera and also an audio recording device that was placed between the two middle members of the groups. Each device (one per group) had two attached directional recording microphones and an additional two directional microphones that were wired and placed facing inwards from the outside of the first and fourth groups members at each table to best capture group conversation. After collection, each audio file was merged with the corresponding video file (as the video recording audio was not of sufficient quality to clearly identify conversation this was removed) to allow individuals and nuances in the conversational recording to be identified. Then each file was cut down to only include conversation from the first to the last talk turn of science discussion related to the question sets. This was important as the audio/video recorders needed to be started at different times based on the needs of the
instructor. Also there was variance in the time instructors notified the recording personnel that the activity was finished, which at times left large sections of post-activity recording in the files.

After cutting excess recording, the remaining files were professionally transcribed. Once transcribed, a researcher went through each transcript to do the following: 1) correct any terms incorrectly transcribed (many terms were science specific and not always picked up correctly by transcribers), 2) watch the video to remove any section of conversation picked up by a recording that came from another group, 3) make sure that each talk turn in the transcript was an individual speaker, and 4) check that a speaker’s talk turn was not being divided into multiple talk turns in the transcript – this was particularly important as the “response to content contribution” code requires that a student is not responding to themselves as this is not group level discourse between peers. Once each of these checks were made before the transcripts were considered to be ready for coding.

**Measures**

In addition to coding student talk, all participants answered several questions as part of a pre-study survey. This survey asked for the gender students identified with (male, female, other) and race (White/Caucasian, Latino/a, Asian/Pacific Islander, African/African American, Multiracial/other). The survey also asked the approximate number of undergraduate science courses they have successfully completed (0, 1-2, 3-5, 6-10, or more than 10) and their current class standing (freshman, sophomore, junior, senior, or post baccalaureate). Finally, participants were asked whether they were planning on pursuing a science-related career in the future (yes, no, undecided). Outside of this survey participants also consented to the use of their classroom grades. This allowed for the collection of a pre-study academic performance indicator (their first
course exam which was a week prior to the intervention), an overall course performance indicator (final percentage grade in the course), and a laboratory specific performance indicator (final grade in their laboratory class).

Participants

Participants were all undergraduate students enrolled in an introductory biology laboratory class and consented to both have their academic scores used and group work audio/video recorded. 107 unique students ($n_{interdependency} = 70$, $n_{comparison} = 37$) were present in at least one recorded groups out of the two data collections. Differences in the number of unique students by condition largely resulted from the difference in the number of classes assigned to each condition. Distribution of gender by condition was relatively consistent with students being 75% female in the interdependency condition and 78% female in the comparison condition. Distribution of race by condition was less consistent with 69% of interdependency condition students identifying themselves as Caucasian/White versus 83% doing so in the comparison condition. This difference can be primarily accounted for by a difference in Latino/a and African American students that together accounted for 20% of students in the interdependency condition, but only 11% in the comparison condition.

Most students had already completed other science courses at the undergraduate level despite the course being part of an introductory biology sequence. The largest percentage of participating students had taken and successfully completed three to five prior undergraduate science courses (46% of participants in each condition). With a smaller percentage having not taken any prior undergraduate science courses (14% of participants in each condition) and only one participant in each condition having had significant experience in undergraduate science
which was indicated by having completed greater than ten prior courses. Similarly, the largest percentage of students in both conditions were at a sophomore level (55% interdependency, 42% comparison), but there were proportionally more students at the junior level in the comparison condition (32%) than in the interdependency condition (14%). A final piece of data collected from students was their interest in pursuing a science career. Approximately the same percentage of students in both conditions (3%) were not planning on pursuing a science career. The remainder showed some variation by condition with proportionally more students being undecided about pursuing a science career in the comparison sample (11%) than in the interdependency condition (3%). In total, the sample of students in the interdependency condition was more racially diverse and at an earlier point in their undergraduate program, but slightly more confident in pursuing a science career in the future. Both samples included students with approximately the same category of previously completed undergraduate science courses and similar gender composition.

**Discourse Coding**

Note that the codes developed for this study were not designed to target differences in student discussion of a specific content area unlike others (e.g. differences in conceptual discussion of evolution- Asterhan & Schwartz, 2009; prediction of fall trajectories – Howe & Zachariou, 2017) so no codes were designed to assess a particular content understanding. Instead we used non-content specific codes that could capture more general discourse moves that either reflected engagement or had prior links to learning. In developing the scheme, we focused on different qualities of student discourse at the talk turn level. A talk turn was defined as an uninterrupted instance of talk by an individual student in the group (similar to Barron, 2003).
We did not focus on affective qualities, like tone or confidence of delivery, nor did we take into account the orientation of speakers in space relative to the group. Instead we focused on breaking down discourse talk turns into basic aspects that have previously been linked to learning. Examining basic properties of discourse at the talk turn level, sometimes referred to as microanalysis or microgenic of talk, is a method that has been previously used to examine discourse differences in relation to learning (Barron, 2003; Chiu, 2008; Howe & Zachariou, 2017). In addition to previous use, we felt that a micro approach to examining discourse would have substantial reliability in coding (as seen in Chiu, 2008) and have high interpretability by readers allowing for replicability through increased ease of use. We did not include codes targeting more emergent or macro properties of conversation that manifest as a larger component of discourse over a number of talk turns or a conversation as a whole. These properties can vary widely in nature with some examples being the relative amount of engagement by all individuals (Xu & Talanquer, 2012), level of cognitive processing (Christian & Talanquer, 2012), and overall nature of group level idea negotiation (Mercer, Wegerif, & Dawes, 1999). This choice was made based on our desire to maximize interpretability and replicability of the coding specifically at the talk turn level.

The final coding scheme went through two rounds of coding and comparison in order to investigate the reliability of its application prior to use. Codes maintained from an initial version, which was piloted on a smaller data set (twelve groups, total talk turns = 514 from another data collection), included on task, content accuracy, justification, and student responses to the content contributions of their peers. These were coded independently by three coders with two coding initially and their agreed upon codes were then compared to a third independent coder. Final inter-rater reliability was good for on task ($k = .97$), accuracy ($k = .87$) and response to content
contribution \((k = .84)\). Justification was surprisingly rare in this sample with only 21 talk turns displaying it. This limited the running of reliability statistics, but raters did correspond in 76% of cases. After this initial reliability check, we decided to expand on the available codes to better account for the alternative factors that may be contributing to learning through discourse. Codes were added to account for if content knowledge was present, input type, and source of a contribution. To get an initial sense of whether the coding scheme was interpretable and useable, three novel coders (2 with Ph.D.’s in science education and one science education graduate student) joined the third coder from the initial coding scheme to code a small sample of 75 talk turns for each code. Feedback from, and discussion with, the novel coders was used to clarify code definitions and resulted in propositions to add subcategories for different form of disagreement and justification. These were not maintained in the final coding scheme as they either occurred too infrequently in the full data set or later were determined to be beyond the scope of the current study respectively. The final coding scheme that emerged from these two rounds of coding and feedback was then applied to all 3,628 talk turns in the final data set for the current study by the primary author. A second rater independently coded 1,185 talk turns (approximately 32% of the total talk turns).

The final coding scheme included seven coding categories (Table 1). Some of these were binary in nature. For example if a talk turn is relevant to the learning activity (on task) or not (off task), if a talk turn contains science content knowledge or not, or if a student provides a justification (via explanation) for their idea or not. We also coded the input type, the source of the contribution relative to the ongoing discussion stream, and accuracy relative to current science understanding (for more details regarding these codes see Table 2 and Supplemental Table 1). Finally we coded how students were responding to the previous talk turn to get a sense
of feedback (mostly agreement or disagreement) in the discussion. The process of coding itself included multiple steps with some codes filtering which talk turns were coded for subsequent codes. This served the dual role of making sure that codes made sense in the larger design of the scheme (e.g. coding justification for statements not on task or coding content accuracy for a talk turn that does not have content presence were not useful) and made coding the entire data set more manageable. For a streamlined version of the coding scheme see Table 2. In this table, black cells represent instances where coding did not take place due to this filtering process. For a more detailed description of each code and of the filtering process see Supplemental Table 1 in Appendix 2.
Table 1
Alignment between mechanisms of learning and discourse codes with examples of studies investigating similar aspects of group interaction (though not always relative to learning)

<table>
<thead>
<tr>
<th>Mechanism</th>
<th>Discourse Codes</th>
<th>Similar aspects of group interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opportunities for information retrieval</td>
<td>• Unique contributions&lt;br&gt;• Reiteration of a contribution</td>
<td>• <em>Amount of talk</em> (Barron, 2003)&lt;br&gt;• <em>Time developing a solution</em> (Canham, Wiley, &amp; Mayer, 2012)&lt;br&gt;• <em>Amount of unique contributions</em> (Samson &amp; Clark, 2011)&lt;br&gt;• <em>Opportunities for group retrieval</em> (Zu, Monsell, &amp; Rubello, 2019)&lt;br&gt;• <em>Repeated group recall</em> (Blumen &amp; Rajaram, 2008)&lt;br&gt;• <em>Repeated group retrieval</em> (Congleton &amp; Rajaram, 2011)&lt;br&gt;• <em>Repetition</em> (Asterhan &amp; Schwartz, 2009)</td>
</tr>
<tr>
<td>Increased accuracy</td>
<td>• Accurate contributions&lt;br&gt;• Inaccurate contributions</td>
<td>• <em>Accurate contributions</em> (Chiu, 2008a, Sampson, Grooms, &amp; Walker, 2010)&lt;br&gt;• <em>Unique accurate contributions</em> (Chiu, 2008b)&lt;br&gt;• <em>Noticing of correct ideas</em> (Barron, 2000)&lt;br&gt;• <em>Inaccurate ideas</em> (Sampson, Grooms, &amp; Walker, 2010)&lt;br&gt;• <em>Misinformation</em> (Christ, Chiu, &amp; Wang, 2014)</td>
</tr>
<tr>
<td>Idea comparison</td>
<td>• Disagreement&lt;br&gt;• Agreement</td>
<td>• <em>Error correction</em> (Ross et al., 2004)&lt;br&gt;• <em>Acceptance and rejection of ideas</em> (Barron, 2003; Sampson &amp; Clark, 2011; Sampson, Grooms, &amp; Walker, 2010)&lt;br&gt;• <em>Rude and polite disagreement</em> (Chiu, 2008b)&lt;br&gt;• <em>Disagreement</em> (Barron, 2003; Purzer, 2011)&lt;br&gt;• <em>Argumentation</em> (Asterhan &amp; Schwartz, 2009; Evagorou &amp; Osborne, 2013)&lt;br&gt;• <em>Supportive and Oppositional statements</em> (Sampson, Grooms, &amp; Walker, 2010)&lt;br&gt;• <em>Agreement and Challenge</em> (Asterhan &amp; Schwartz, 2009)&lt;br&gt;• <em>Identifies/Corrects misinformation and agrees</em> (Christ, Chiu, &amp; Wang, 2014)</td>
</tr>
<tr>
<td>Justification of ideas through explanation</td>
<td>• Justified contributions</td>
<td>• <em>Justification via evidence</em> (Clark &amp; Sampson, 2008)&lt;br&gt;• <em>Justification of ideas</em> (Chiu, 2008a; Sampson &amp; Clark, 2011)&lt;br&gt;• <em>Rebuttal</em> (Asterhan &amp; Schwartz, 2009)&lt;br&gt;• <em>Explanation</em> (Hausmann, Chi, &amp; Roy, 2004; Webb, 1991)</td>
</tr>
</tbody>
</table>
Table 2
Final coding scheme with embedded example.

<table>
<thead>
<tr>
<th>ID</th>
<th>Talk Turn</th>
<th>On task</th>
<th>Content present</th>
<th>Input type</th>
<th>Contribution Source</th>
<th>Content Accuracy</th>
<th>Justification</th>
<th>RCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>maybe take your... Maybe take them out of the ground, I don't know</td>
<td>1</td>
<td>1</td>
<td>S</td>
<td>U</td>
<td>N</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>118</td>
<td>maybe move to a different location</td>
<td>1</td>
<td>1</td>
<td>S</td>
<td>U</td>
<td>A</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>69</td>
<td>would adding more bacteria help?</td>
<td>1</td>
<td>1</td>
<td>QS</td>
<td>U</td>
<td>IA</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>118</td>
<td>yeah maybe to combat</td>
<td>1</td>
<td>1</td>
<td>S</td>
<td>U</td>
<td>IA</td>
<td>0</td>
<td>A</td>
</tr>
<tr>
<td>5</td>
<td>wouldn't that just be a cycle though where you end up with more phage</td>
<td>1</td>
<td>1</td>
<td>S</td>
<td>U</td>
<td>B</td>
<td>0</td>
<td>D</td>
</tr>
<tr>
<td>118</td>
<td>more phage replicating [nodding head in agreement]</td>
<td>1</td>
<td>1</td>
<td>S</td>
<td>R</td>
<td>B</td>
<td>0</td>
<td>A</td>
</tr>
<tr>
<td>69</td>
<td>what if when you are making the wine you do it in a new area. Because maybe the temperature is too high</td>
<td>1</td>
<td>1</td>
<td>S</td>
<td>U</td>
<td>A</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>5</td>
<td>oh that’s good. temperature. temperature.</td>
<td>1</td>
<td>0</td>
<td>S</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>118</td>
<td>a temperature controlled environment.</td>
<td>1</td>
<td>1</td>
<td>S</td>
<td>U</td>
<td>A</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>5</td>
<td>boom</td>
<td>1</td>
<td>0</td>
<td>S</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>118</td>
<td>Done!</td>
<td>1</td>
<td>0</td>
<td>S</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 On task. Includes yes [1] and no [0]. Separates talk turns that are in alignment with the learning activity from those that are not. On task talk turns include contributions, questions, logistic statements, and other statements that are oriented towards completing the task. Off task talk turns are those that have no relationship to completing the task or task material in any way.

2 Content present. Includes yes [1] and no [0]. Includes any statement demonstrating science knowledge or requires an understanding of a science principle, idea, or terminology. Note that the connection to content must be explicitly present during the talk turn.

3 Input type. Includes statements [S], questions [Q], and question statements [QS]. This separates talk turns that are put forward as a question by the speaker from those that are statements. Additionally, a question statement is an utterance posed as a question but is functioning as a statement.

4 Contribution source. Includes unique content contribution [C], reiterated contribution [R], information directly from the provided task sheet/information [T], or coming directly from interaction with the instructor [I]. Primarily used to determine the uniqueness of ideas in the group and remove ideas from sources (task materials and instructor) outside of the group itself prior to analysis. Only coded if content is present.

5 Content Accuracy. Includes accurate [A], inaccurate [IA], combination of accurate and inaccurate [ / ], and not enough information present [N]. The purpose of this coding category is to provide information on the accuracy of content utterances by students. Only coded if content is present.

6 Justification. Includes a justification (explanation as to why) of one’s contribution [1] or no justification being present [0]. Only coded if content is present.

7 Response to content contribution. Includes explicit agreement [EA], disagreement [D], both agree and disagree [B], neither explicitly agree nor disagree [N]. Primarily used to track how students respond to the contributions of peers in conversation. Only coded if content is present.
Analyses

In order to examine differences in student performance by condition, parametric analysis were most appropriate given the relative fitting of student quiz scores to a normal distribution. For this analysis multiple regression was used. Multiple regression allowed overall student performance in the laboratory class to be entered into the model prior to conditional effects and thus control for some variance in overall student ability and knowledge of class material. Reported results include unstandardized $\beta$ values which are more easily translatable to student grade performance than standardized $\beta$ values. The adjusted $R^2$ values are also reported as these indicate the amount of variance accounted for by the overall model (i.e. prediction of student scores) with larger values represent a greater ability of a predictor to account for this variance.

The distribution of discourse codes were examined for normality using a Shapiro-Wilk $W$ Test. Results revealed that all discourse count distributions showed significant deviation from a normal distribution making parametric analysis inappropriate (e.g. overall talk turns, $W = .59, p < .0001$). This is common for dependent variables that are count data because discrete count data includes rare events that violate even approximate normality (Myers et al., 2012). This non-normality, the use of count variables (number of discourse incidences present), and visual examination of distributions supported the use of a generalized linear model for regression. Generalized linear models have three major components; (a) specified response distribution (based on an observed distribution), (b) a linear predictor (in this case group presence in different conditions or identification as higher/lower performing group), and (c) a link function that functions as a transformation connecting the two (Myers et al., 2012). One of the most common is Poisson regression, yet even count data can violate the expectations of the Poisson regression
model if the variance is too large relative to the mean. If this is the case, then a negative binomial model or dispersion parameter can better fit the collected data (Gardner, Mulvey, & Shaw, 1995). Yet both of these models tend to misrepresent the data if there are many zeros in the data set. When the data contains excess zeroes, there are both zero inflated (ZI) versions of Poisson and Negative binomial models that can be used to account for this discrepancy (Cheung, 2002).

Given the variety of code distributions in this study, each generalized linear model had to be specified based on the distribution that best represented that of each discourse code. Comparative Akaike information criterion (AIC) values were used to determine optimal model fit between the prior discussed models. Lower AIC values represent less information loss and indicate a better fitting model (Anderson, Burnham, & White, 1998). All generalized linear models were examined for each potential distribution (Poisson, ZI Poisson, Negative Binomial, and ZI Negative Binomial) and the distribution which produced the lowest AIC value was adopted and reported in the results.

All analysis was completed using JMP PRO 12.2.0. Generalized regression was used to examine if condition predicted differences in behavioral amounts. The resulting R-squared values reported are Craig and Uhler generalized R-squared values, which are a normalized version of Cox and Snell’s pseudo R-squared value allowing for interpretation similar to that of traditional linear R-squared values (Nagelkerke, 1991).

Accounting for alternative factors in analysis

It is important to note that there were strong relationships between the total amount of content related talk a group engaged in and the frequency of their discourse codes. This is
unsurprising given that each additional talk turn provides an opportunity for students to exhibit one of the discourse codes. The lowest correlation between content talk length and a discourse code was $\rho = .63$ (justification) and the highest was $\rho = .96$ (unique contribution) with the majority of the remainder falling between .60-.80. These extremely high correlations suggested that if total content talk turns was not statically controlled for during analysis, then observable differences between condition or group performance level could be resulting from differences in group discussion length and not the frequency of the discourse move itself. Thus for all analysis examining differences in discourse moves for research questions two and three we always entered the number of content talk turns into the model first to statistically control for this variance prior to entering the predictor of interest (e.g. condition or group performance level).

There was also the potential for differences in both conditional comparisons as well as those between performance groups to have been driven by a third variable (general academic ability/content knowledge). To account for this, we identified two proxies for this factor-students’ score on their first exam (a week before the intervention) and student grade in the course overall (laboratory and lecture components combined). Both of these measures were standard across all students regardless of condition or performance group membership as they were all graded with the same criteria. Non-significant difference in these factors were established prior to analyzing differences in performance by condition (see below for details). Both overall performance, as well as which quiz they were taking (out of 2), were entered into the regression model prior to examining the ability of condition presence to predict differences in student performance. The finding of a non-significant difference in student academic ability between conditions should apply to discourse comparisons by condition, but to check we also entered average group performance (in the course overall) into the regression model prior to
adding condition as a factor. Average group performance in the course overall did not predict differences in group discourse in 6/7 categories. Only instances of disagreement were predicted below a typical alpha value of $p < .05$ (value of $p = .04$), but this was larger than the alpha value used in this study ($p < .007$) when correcting for multiple comparisons. These results largely support that overall student ability in the course was not driving conditional difference in group discourse.

This same process, of examining general student academic ability in the course overall prior to discourse differences, was also followed for comparing the highest and lowest performing groups. Results showed that the highest performing groups had significantly higher average overall course performance than the lowest performing groups ($m_{\text{highest}} = 83.5$, SE = 2.2; $m_{\text{lowest}} = 77$, SE = 2.2, $p = .03$, $R^2 = .10$). To control for this difference we added average group performance in the course overall into the regression model to control for this variance prior to using highest/lowest group performance membership (categorical) as a predictor of each discourse move. Despite overall course performance being significantly different between the highest/lowest performing groups, in none of the discourse comparisons was average group performance a significant predictor of discourse differences. Thus in the final models this factor was removed as it was not able to control for significant variance in discourse and resulted in worse overall model fit (evaluated by AIC).

**Discourse codes used for analysis**

Comparisons of discourse were restricted to the codes that had one of the four theoretical links to being able to support or detract from group learning processes (see the “Learning through discourse” section above). First, codes related to opportunities for information retrieval
were included. Each time an idea is put forth to the group this presents a recall opportunity for group members. This can manifest through either increased unique opportunities for recall (unique contribution code) or through more opportunities to recall a particular contributed idea through that idea coming up multiple times during conversation (reiterated contribution code). Thus both the unique and reiterated contribution codes were included for comparison.

Second, codes addressing content accuracy in groups were included (accurate contribution and inaccurate contribution). This is because increased accuracy of a conversation means that the process of conversation is more likely to reinforce accurate cognitive knowledge structures which students can draw on at a later point (e.g. when completing an achievement measure). In an opposing respect, subsequent achievement would suffer if groups are discussing content inaccurately and these inaccurate ideas are being reinforced through discourse (given no error correcting mechanism). Thus both the number of accurate contributions and inaccurate contributions were compared between conditions, but more ambiguous codes, like contributions that have both accurate and inaccurate aspects as well as those with not enough information to judge accuracy, were not because the clarity of their connection to potential learning is more opaque.

Third, codes that requires students to provide feedback to their group members and thus instances of idea comparison were included. The process of providing feedback requires that students actively check their own understanding against that of peers. This process has the potential to both formalize one’s understanding (i.e. make it explicit) and push that of other group members. Idea comparison was captured in both the agreement and disagreement codes and used for comparison purposes in analysis. Both simultaneous “agreement and disagreement”
as well as different forms of disagreement were not present in high enough frequencies (often only 0-1 times a group) for analyses.

Fourth, when students justify their ideas this results in an explanation that can reinforce a student’s existing knowledge structures through the explicit formalizing and communication of their ideas. The process of justification can manifest through spontaneous choice by a student (not pushed for by other groups members) when presenting an idea or can be catalyzed by through disagreement by others in the groups. In either case justification may promote learning and thus was the final code used for comparison.

In total, the inclusion of the aforementioned codes resulted in seven comparisons for analyses examining differences by condition and group performance. Given that each analysis involved multiple comparisons we needed to adjust our alpha value according to reduce the impact of false positive results. Thus each type of analysis (conditional effects or performance differences) was considered a family and with a Bonferroni correction (.05/7) this brought our alpha value to less than .007. Thus only results lower than this adjusted alpha were considered statistically significant for the purposes of this study.

Results

Research Question (1): Did students perform significantly better in the interdependency condition?

In order to assess the impact of interdependent structuring on student achievement in the laboratory classrooms, each student completed a quiz designed to measure their understanding of the concepts from the session a week after completion. Prior to examining conditional differences in outcomes, potential differences in student academic ability for the sampled students was examined for each condition. Indicators of academic ability included both their
performance in the course overall as well as their first exam score (taken the week prior to the first data collection). T-test results showed that there were non-significant differences in both student’s first exam scores ($m_{\text{inter}} = 69, \ SE_{\text{inter}} = 2; m_{\text{comp}} = 71, \ SE_{\text{comp}} = 3, \ p > .60$) and overall performance in the course ($m_{\text{inter}} = 82, \ SE_{\text{inter}} = 2; m_{\text{comp}} = 82, \ SE_{\text{comp}} = 2, \ p > .80$) between students in each condition. These results suggest that the sample of students from each condition did not significantly differ in their academic ability relative to the course.

Next conditional differences in the outcome quizzes were examined. Using multiple regression, condition was used to predict differences in individual student achievement when statically controlling for their overall performance and which quiz they were taking (2 total). Presence in the interdependency condition predicted a significant increase in score ($p < .05$) on the subsequent week’s quiz ($n = 138, m = 1.83, \ SE = .03$) than did presence in the comparison condition ($n = 74, m = 1.74, \ SE = .04$). This difference was relatively minor in effect ($R^2 = .04$), and equated to a 5 percentage point increase on average for interdependency participation (Figure 1). This result supports that students in the interdependency condition showed significantly higher achievement in the subsequent week’s quiz.
Research Question (2): How did interdependent structuring shift group discourse?

Next characteristics of group discourse were compared between groups in the interdependency condition to those in the comparison condition. Specifically, group presence in Phase 2 were contrasted to those in the comparison condition. This was because both comparison and Phase 2 groups were completing all of the questions for the session (as opposed to in Phase 1 of the interdependency condition where they were only completing one question). Thus restricting this analysis to only Phase 2 groups reduced the likelihood of either the specific question they are answering (3 per session and they varied in both difficulty and content) or the amount of questions of driving significant effects. In addition, this choice allowed for a more focused comparison. This is because Phase 1 of the intervention primarily serves to establish specialized student knowledge which then theoretically should manifest as differences in the social atmosphere in which students engage during Phase 2. Thus a comparison of Phase 2 discourse to that of the control was used directly below.

Figure 1: Mean quiz performance by condition. Student in the interdependency condition scored significantly higher ($p < .05$). Maximum 2 points in increments of .25 pts.
Results showed that condition predicted significant differences in three of seven discourse codes (reiteration, accuracy, and justification of ideas) after correcting for multiple comparisons ($p < .007$). Specifically, groups in the interdependency condition spent more time reiterating ideas in their group than did the comparison when statistically controlling for the total number of content talk turns of the group ($b = .65, SE = .15, p < .0001$, condition $R^2 = .05$, overall $R^2 = .88$). Groups in the interdependency condition also had more accurate ideas in general ($b = .38, SE = .07, p < .0001$, condition $R^2 = .05$, overall $R^2 = .86$) and engaged in significantly more justification of their ideas ($b = .78, SE = .08, p < .0001$, condition $R^2 = .19$, overall $R^2 = .63$). Together the results show that interdependency groups engaged in three discourse codes (reiteration of ideas, presentation of accurate ideas, and justification of ideas) at higher rates than in the comparison (when controlling for total content talk turns). No significant difference was seen for the remaining discourse codes. For a graphical representation of how justification varied between conditions (when controlling for total content talk turns) see Figure 2.
In order to provide some insight into the relative frequencies of discourse codes (including their variance between conditions) we used the regression equations to predict the amount of each discourse code that would be present in a typical group engaging in 20 talk turns of discourse. Results can be seen in Figure 3. To generate this figure we specified that the group would have engaged in twenty talk turns (mean of the overall sample). Each column represents the predicted number of talk turns that would include a specific discourse code out of the twenty with a 95% confidence interval. It should be noted that the same talk turn can exhibit multiple codes (e.g. a contribution can simultaneously be unique, accurate, and justified).

Figure 2: Difference in the relationship between group use of justification and total content talk turns in the group by condition. Difference between conditions was significant ($b = .78$, $SE = .08$, $p < .0001$). Condition $R^2 = .19$, Overall model $R^2 = .63$. A generalized linear model was used. This had a specified Poisson distribution with overdispersion and a log link function as this was the optimal model based on relative fit comparisons (AIC).
Next we examined the extent to which the phase structure could account for variation in student discourse moves seen between conditions. This was evaluated by looking at differences in student discourse between Phase 1 and Phase 2 in the intervention only. We used phase as a predictor for each of the discourse differences seen between conditions above (reiteration, accuracy, and justification of ideas) with total content talk turns first being entered into the model to control for difference in group talk length.

We hypothesized that the phase structure could result in increased accuracy between phases with Phase 1 acting as an initial filter of ideas within groups so that only those ideas deemed as most accurate by the group would be carried into Phase 2. We also hypothesized that

* p < .007 (alpha value after Bonferroni correction)

Figure 3: Predicted differences in discourse codes between interdependency (Phase 2, n = 25) and comparison (n = 28) groups. Each set of bars made use of the best fitting model per discourse code (determined by relative AIC) to predict the number of each discourse code that would be seen in a group with 20 content talk turns. Error bars represent the 95% confidence interval for the predicted value.
justification of ideas would be highest during Phase 2. This is because interdependency should be increasing the level of social responsibility of students in Phase 2 as their peers are depending and expecting an individual to bring with them accurate and insightful content relative to their specific question from Phase 1. This means that the ability of a student to not participate is very limited by the structure of the task. In this scenario students would want to be better prepared for Phase 2 than under normal group question completion and have relatively clear explanations and justifications for their answer to provide to their peers. Also students would be expected to have greater confidence in justifying their ideas in front of peers following Phase 1 given that they have had a chance to discuss potential explanations and come to those best justified. Finally, we hypothesized that reiteration of ideas could actually either increase or decrease between Phases 1 and 2 based on the function of this reiteration. For example, Phase 2 would be expected to have more reiteration of ideas given that students are hearing ideas for the first time and may need to clarify points or even just ask them to be repeated for either recording or review purposes. In contrast, given the social expectation of Phase 2, students may spend more time reviewing ideas (which could increases reiteration) in their preparation for Phase 2. In this case we would see a decrease in reiteration between phases.

Results showed that reiteration, accuracy, and justification of ideas all significantly increased between Phases 1 and 2, but to different degrees (Figure 4). Specifically, group presence in a Phase 2 predicted significantly more reiteration of ideas ($b = .41, SE = .14, p < .005$, condition $R^2 = .04$, overall $R^2 = .59$), accuracy of ideas ($b = .23, SE = .07, p < .001$, condition $R^2 = .02$, overall $R^2 = .72$), and justification of ideas ($b = .64, SE = .12, p < .0001$, condition $R^2 = .14$, overall $R^2 = .50$) when statistically controlling for variance in content talk turn amounts. These results support that overall differences in discourse between conditions are localized to Phase 2.
As a final check to assess that the effects of the intervention are localized to the Phase 2 portion of the intervention (as hypothesized), we examined the shifts in discourse moves between the Phase 1 intervention groups and comparison groups. Given that interdependency is not strongly present until Phase 2, when students have specialized knowledge needed by their peers, we hypothesized that there should be minimal shifts in discourse between these groups. If shifts were seen between Phase 1 intervention and comparison groups we predicted that these would most likely result from comparison groups having had to answer three questions while Phase 1 intervention groups only answered one. Amount of content talk turns was added first to control for talk turn length when we built the models for this comparison. Results showed no significant

*Figure 4: Predicted differences in discourse codes between groups at different phases of the interdependency intervention (n = 25 per phase). Each set of bars made use of the best fitting model per discourse code (determined by relative AIC) to predict the number of each discourse code that would be seen in a group with 20 content talk turns. Error bars represent the 95% confidence interval for the predicted value. Only codes found to be significantly different between conditions were used to confirm that increases were occurring between phases.

*p < .016 (alpha value after Bonferroni correction)
differences (either $p < .05$ or $p < .007$) were found in group discourse between the Phase 1 interdependency groups and comparison groups. This provides further support that the shifts in student discourse were localized to Phase 2 (as hypothesized) and thus shifts in discourse during Phase 2 may have had a substantial role in driving the differences in learning seen in the result from research question #1.

Research Question (3): How does discourse vary between the highest (4th quartile) and lowest (1st quartile) performing student groups?

Examining the difference in discourse between higher and lower performing groups have repeatedly been used to explore what forms of talk may promote learning (Barron, 2000; Barron, 2003; Evagorou & Osborne, 2013; Khosa & Volet, 2014; Sampson & Clark, 2011) as they provide explicit comparisons that lend themselves to contrasting student discourse patterns. We originally defined higher performing groups as those that scored at least a 90% of better on average on the outcome quiz. Yet this separation point was based on an absolute criteria and the fact that it put approximately half of the groups into the “high performance” group indicated that the actual distribution of the group scores on the outcome measure were higher than what might normally be seen. Thus we decided to instead separate groups into relative levels of performance through quartiles. This resulted in much more distinct groups with 20 groups being categorized as those with the lowest performance (1st quartile, $m_{group} = 78\%$) and 20 groups being categorized as those with the highest performance (4th quartile, $m_{group} = 99\%$). As noted above, this separation resulted in groups that were significantly different in terms of their average overall course performance. Differences in course academic ability could be driving any potential differences seen in discourse moves between groups, so average overall performance (group level) was first
added into each regression model to assess its potential contribution to discourse differences. Average overall course performance did not significantly predict differences in group discourse for any code and was removed from the model prior to adding group performance membership (highest or lowest) into the model.

Results showed that group presence in the higher performing category predicted significant differences in three of seven discourse codes (accuracy of contribution, disagreement, and justification) after correcting for multiple comparisons (p <.007). Higher preforming groups spent significantly more time contributing accurate ideas after statistically controlling for the total number of group talk turns ($b = .52$, $SE = .21$, $p <.005$, performance $R^2 = .13$, overall $R^2 = .70$). Higher performing groups also engaged in significantly more disagreement about ideas ($b = .90$, $SE = .33$, $p <.007$, performance $R^2 = .09$, overall $R^2 = .67$) and provided significantly more justification of their ideas ($b = .53$, $SE = .16$, $p <.005$, performance $R^2 = .26$, overall $R^2 = .62$). Together these results suggest that higher performing groups were those that are able to present more accurate ideas during conversation, were more willing to disagree with presented ideas, and justify their ideas during discussion. For a graphical presentation of differences in discourse between higher and lower groups see Figure 5. This was generated in the same method discussed above.
The results of the research questions, when taken together, provide some insight into potential contributions that discourse moves (as driven by the interdependency intervention) may have contributed to student learning. The results of research question one (Did students perform significantly better in the interdependency condition?) showed that student presence in an interdependency condition predicted greater performance beyond what could be accounted for by their performance in the course overall. Given that the experimental manipulation was a relatively lightweight restructuring of how students completed the activity there are limited potential mechanisms through which this difference may have occurred. Students’ collaborative engagement shifts in response to interdependency (Brewer & Klein, 2006; Buchs et al., 2004; Jensen, Johnson, & Johnson, 2002; Lew et al., 1986; Moser & Wodzicki, 2007; Premo &

**Figure 5:** Predicted differences in discourse codes between the highest and lowest quartile of performing groups (n = 20 per each). Each set of bars made use of the best fitting model per discourse code (determined by relative AIC) to predict the number of each discourse code that would be seen in a group with 20 content talk turns. Error bars represent the 95% confidence interval for the predicted value.

*p < .007 (alpha value after Bonferroni correction)*

**Conclusion**

The results of the research questions, when taken together, provide some insight into potential contributions that discourse moves (as driven by the interdependency intervention) may have contributed to student learning. The results of research question one (Did students perform significantly better in the interdependency condition?) showed that student presence in an interdependency condition predicted greater performance beyond what could be accounted for by their performance in the course overall. Given that the experimental manipulation was a relatively lightweight restructuring of how students completed the activity there are limited potential mechanisms through which this difference may have occurred. Students’ collaborative engagement shifts in response to interdependency (Brewer & Klein, 2006; Buchs et al., 2004; Jensen, Johnson, & Johnson, 2002; Lew et al., 1986; Moser & Wodzicki, 2007; Premo &
Cavagnetto, 2018), but a deeper discourse analysis was needed to examine if this increased collaborative engagement may related to mechanisms of group learning in the current study.

To answer research question two (How did interdependent structuring shift group discourse?) we coded different aspects of discourse previously linked to learning to see the extent to which shifts occurred in response to the intervention. The results showed that rates for three aspects of discourse (reiteration, accurate, and justified contributions) related to three different potential mechanisms of learning through dialog (opportunities for information retrieval, increased idea accuracy, and justification of ideas respectively) were significantly increased within interdependency groups. These aligned to some of the differences in discourse between the highest and lowest performing groups in research question three (How does discourse vary between the highest (4th quartile) and lowest (1st quartile) performing student groups?). Specifically, two of the shifts in group discourse (accurate and justified contributions) were in common with groups in the interdependency intervention. This finding suggests that both increased accuracy by groups and higher rates of idea justification may play a role in the significantly increased performance in the interdependency condition. It should also be noted that idea comparison (specifically disagreement and not agreement) was also found to have significantly increased prevalence in the highest performing groups, but no conditional shifts were seen. This suggests that disagreement may play a role in group learning, but is not promoted by interdependent structuring, and thus unlikely to be systematically contributing to the learning gains seen in the experimental condition.

Perhaps more important than the statistical significance of these shifts in discourse is the relative effect that each is having on overall group discourse. To assess this factor we reported generalized r-square values to inform the percentage of overall variance in each discourse code
which was accounted for. Interpreting these values provides insight into which discourse shifts were most dramatic and thus may have a relatively larger impact on group learning. For an overview of this comparison see Table 3 below. Of particular note is the difference in the magnitude of shifts in justification ($R^2 = .19$, large effect) which was almost four times that of either reiteration of contributions ($R^2 = .05$, small effect) and accurate contributions ($R^2 = .05$, small effect). This demonstrates the primacy of the interdependency task in pushing for increased student justification of their ideas during group discourse. Justification also showed the most substantial difference between the highest and lowest quartile of group performance ($R^2 = .26$, large effect) with both accurate contributions ($R^2 = .13$, medium effect) and disagreement ($R^2 = .09$, medium effect) have a smaller effect. In total, both condition and level of group performance saw the largest magnitude of differences in discourse manifest as increased student justification. This finding may indicate that student justification of their ideas, through providing an explanation of “why”, may be a key discourse practice driving student learning in the current study.

Table 3.  
Magnitude and alignment of discourse shifts

<table>
<thead>
<tr>
<th>Discourse Code</th>
<th>Condition$^1$</th>
<th>Performance$^2$</th>
<th>Alignment$^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reiteration of a contribution</td>
<td>5 (small)</td>
<td>ns</td>
<td>none</td>
</tr>
<tr>
<td>Accurate contribution</td>
<td>5 (small)</td>
<td>13 (medium)</td>
<td>some</td>
</tr>
<tr>
<td>Justified contribution</td>
<td>19 (large)</td>
<td>26 (large)</td>
<td>substantial</td>
</tr>
<tr>
<td>Disagreement</td>
<td>ns</td>
<td>9 (medium)</td>
<td>none</td>
</tr>
</tbody>
</table>

Note: Values represent the percentage of variance explained followed by the relative effect size based on Cohen (1988).  
$ns$ = non-significant difference  
$^1$ group presence in the interdependency condition  
$^2$ highest performing quartile  
$^3$ based on whether a significant shift is seen in both experimental and highest performing groups (some or none) and these are considered substantial if the shift is within the same effect size category
Discussion

The impetus for the current study was to gain greater clarity to the potential mechanisms by which structuring learning to promote interdependency may shift student group discourse in ways that promote learning. This desires stemmed from the longstanding need for instructors to know the features of group discourse critical to learning in order to encourage these in their classrooms (Webb, 1991). While research on student discourse has advanced over the last few decades few studies have addressed peer-peer dialog as a mechanism of learning (Howe, 2013). In fact a review of 225 studies looking at classroom dialogue conducted by Howe & Abedin (2013) found that only 15 related classroom dialogue to achievement with only 9 of these examining group discourse. More recent commentary also recognizes the need for greater research to understand how students learn from others in groups (Kuhn, 2015) including the role of discourse as a mechanism of learning (Howe, 2017). It is in this context that the current study contributes. Findings reinforce that discourse can systematically varies between higher and lower performing groups, but be independent of overall academic ability of the participants (Barron, 2003). It is not academic ability, but how groups engage in discourse, that may be ultimately key to successful learning.

Not all mechanisms of learning were evidenced in the current study. Notable is the finding that increased opportunities for information retrieval (i.e. through increased unique contributions or reiteration of contributions) was not a characteristic of high achieving groups. Though the intervention did significantly increase the reiteration of contributions, not having this in common with high performing groups suggests that this is not contributing to greater learning through interdependent structuring. In contrast, disagreement (one aspect of idea comparison)
was found to be significantly increased in higher performing groups, but was not shifted by interdependent structuring. This suggests that disagreement may be a critical aspect of idea comparison that contributes to learning, but that agreement (which also can involve checking a peer’s idea against your own) does not. This may be due to the ease in which one can verbally agree without needing to check the alignment between the suggested idea and one’s own. While this can also occur with disagreement, results of the current study suggest that disagreement is differentially related to learning than is agreement when a group compares ideas.

Of particular interest is the role that student justification of their ideas may have had in promoting group learning. The rate of groups justifying their ideas showed the most substantial shifts (large effect) of any discourse codes for both condition and group performance. It should be noted that we defined justification as a student providing a rationale or explanation to support their contribution. Being able to justify one’s ideas is a critical aspect of science communication (NRC, 2000; 2012; V&C, 2011) so the ability of interdependent structuring to promote this aspect of discourse may simultaneously be able to promote this skill and encourage greater student learning. Interdependent structuring of tasks is a fairly lightweight design change and is easily adaptable to a variety of learning contexts and activity types making the intervention a straightforward way to push students to practice justifying their ideas while communicating about science content.

While the results of the present study are encouraging, continued research into the relationships between group discourse and learning are needed- especially as these relate to different activity structures. Knowing characteristics of productive discourse are valuable, but cannot be used to increase student learning unless we know how to promote them. Part of this work should be greater investigation into different mechanisms of learning through discourse
across different cooperative activities. While the current study attempts to contribute to this process, there are substantial limitations that should be addressed in future work. The implications of the study are limited by only capturing two different instances of group work (two different weeks) which limits generalizability. This sample was chosen due to the constraints of having to process and coding large amounts of videos, which is likely a rationale for the presence of limited samples for discourse research in general. Also the focus of the current study is on the group and not individual level. While this provides straightforward insight into the general types of discourse occurring in groups it does not provide a distinction between the role discourse can have for a speaker versus listener. Given research findings that the impact of discourse on learning varies between speaker and listener (Abel & Roediger, 2018; Hausmann, Chi, & Roy, 2004; Webb, 1991) this is an additional nuance to the discourse that could have important implications for how interdependent structures could impact learning. Particularly because this structuring promotes each student to enter their Phase 2 group with a unique contribution from their Phase 1 group. This puts students in a position where they have listened to peer ideas in their Phase 1 groups and then have had to internalize these in order to provide them as a speaker in their Phase 2 groups. In this way, even students that only listened in Phase 1 have the same need to contribute as those that actively spoke and participated. Understanding if requiring spoken contributing during Phase 2 is able to increase learning for those only listening in Phase 1 would provide valuable insight into the conditions in which listening and speaking may impact student cognition in groups.

A final, and considerable, limitation of the current study is that it does not address additional variables that may be driving discourse differences at the individual level (motivation, science interest, views of the social environment) or group level (equality of talk,
gender/racial/ability differences in group composition) all of which have the potential to also be contributing to a group’s discourse and thus learning. All of these have been investigated to some extent by the field in terms of their contributions to student experience in groups, but few investigations have quantitatively explored how these factors intersect with observed discourse and learning. Further work looking into how these factors shape observable group discourse and the resultant learning is necessary to understand the contexts in which productive discourse arises- including where greater or lesser supports may be needed.

Despite these limitations, the current study does suggest that interdependent structuring can shift group discourse in ways that support greater student learning. The finding that group performance was related to discourse shifts, but the academic ability of the groups (as indicated by average overall course performance) was not, indicates that discourse was most likely driving group learning over academic ability. An implication of this finding is that the types of discourse present in peer conversations is key aspect of the process by which learn in groups occurs. This should be kept in mind as research continues into the role that student experience with social aspects of science play in science classrooms. Integration of social aspects of science practice into the classroom may play a multifaceted role in shaping student views of science and encouraging development of science related social skills, while simultaneously promoting greater learning.
References


DISSERTATION SUMMARY

The dissertation research line began by investigating the relationships between students’ perceptions of their classroom social environment, perception of benefit from peers, and prosocial dispositions to assess the extent to which these aligned with interdisciplinary research on cooperation (Dissertation Article #1- Premo, Lamb, & Cavagnetto, 2018). This was done using structural equation modeling on CCEM data collected from 845 undergraduate students enrolled in undergraduate biology courses. The study sought to assess models that accounted for either students’ prosocial dispositions or disposition towards enforcing cooperation in the classroom. For the prosocial disposition models, reciprocity and friendship in a classroom were thought to be related to the benefit students perceived as coming from classmates and thus their willingness to expend energy in helping their peers. For the enforcement model it was hypothesized that both social group characteristics (friendship and reciprocity) as well as perceived benefit from classmates would independently predict students’ disposition towards enforcing cooperation. Results showed that the hypothesized models had the best fit to the data based on both relative (AIC) and absolute (RMSEA, CFI, & SRMR) fit indices. This provided empirical support that factors in the classroom social environment were able to predict student willingness to cooperate. This suggested that future studies attempting to promote greater student prosocial dispositions should increase the benefit student perceive as coming from their classmates during collaborative learning experiences in the classroom.

Having established that student perception of benefit coming from their peers predicts prosocial disposition, the subsequent study sought to determine the extent to which interdependent structuring (which should increase the likelihood of beneficial interactions occurring with peers) could increase observable whole-classroom cooperative behavior (Dissertation Article #2- Premo, Cavagnetto, & Davis, 2018). Ten laboratory classrooms of an
introductory sequence biology course experienced an interdependency intervention over nine weeks. An additional ten classrooms completed the same tasks during the same period, but without structured interdependency. Multiple sources of data were collected as outcomes. These included brief weekly whole-classroom observations (CCOP), pre-post collection of student perceptions of their classroom social environment (CCEM), and six weekly quizzes that served as learning outcomes. In addition, 24 students (12 per condition) were audio/video recorded at three different time points and then coded for the ways they engaged in cooperation. This allowed for an examination of differences in collaborative engagement at the individual level.

Results of the large-scale intervention provided insights into the functional role that interdependent structuring may play in promoting whole-classroom peer engagement and learning. In contrast to expectations, classrooms participating in the interdependency intervention did not demonstrate significantly greater achievement on weekly quizzes. Despite this finding, classrooms participating in the intervention did demonstrate significantly higher observed whole-classroom prosocial behavior and engagement with discussion practices \((p < .05)\) and showed significantly increased prosocial behavior over the period of the intervention \((p < .05)\). In addition, observations of individual student engagement showed significantly more whole group science discussion, social support of peers, and significantly less subgroup talk \((p < .001)\). All of these findings suggested that that the intervention was able to promote greater student engagement with peers during science learning. Finally, student perception of their classroom environment (CCEM) were examined to see if shifts occurred in response to the intervention and which of these factors predicted whole-classroom collaborative engagement. While no systematic shifts were seen in individual student perceptions, changes in many different aspects of students’ social environment did predict whole-classroom collaborative engagement outside of
intervention effects. For example, the interaction between changes in reciprocity and student relative investment in cooperation was the strongest factor predicting late whole-classroom prosocial behavior \((p < .05, R^2 = .17)\). A wide variety of aspects of the classroom social environment significantly predicted late intervention discussion practices beyond the effects of the intervention itself. The strongest was a class’s initial perception of peer reciprocity \((p < .05, R^2 = .09)\) with change in reputational concern during the semester predicting decreased whole-class engagement with discussion practices \((p < .05, R^2 = .07)\). Taken together, the results of the study support that the interdependency intervention was able to significantly increase collaborative engagement at both the whole-classroom and individual levels. Results also showed that a variety of additional factors beyond the intervention were able to predict whole-classroom cooperative engagement.

Not seeing a significant increase in student achievement was concerning and called for an in-depth examination of the relationship between how students are engaging with peers through discourse and learning. To address this, the final study in the dissertation research line was designed to specifically examine nuances in group discourse within the intervention, between the intervention and the comparison condition, and between the highest and lowest achieving groups. Video and audio data was collected from 78 student groups \((n_{\text{interdependency}} = 50, n_{\text{comparison}} = 28)\) and each student talk turn (3,628 total talk turns) was coded for each of seven different categories. These categories included those which were engagement related (on/off task, content presence, input type) and those with aspects previously linked to group learning (contribution source, content accuracy, justification, and response to peers’ content contributions).

Results showed that group participation in the interdependency intervention predicted significant increases in the accuracy of student contributions, amount of reiteration of ideas and
the justification of ideas (all $p < .007$, alpha shift due to Bonferroni correction). Comparisons of the highest and lowest performing groups showed that the highest performing groups’ discourse included significantly increased accuracy of student contributions, justification of ideas, and disagreement with peer ideas ($p < .007$). When these results are taken into account, alongside students in the intervention achieving significantly higher scores on the outcome measure ($p < .05$, $R^2 = .04$), this suggests that the intervention may have been able to promote greater group learning through encouraging student accuracy and justification of ideas- as these are in common amongst the intervention groups and those which were the highest performing.

**CONTRIBUTIONS TO THE FIELD**

The three studies found in this dissertation have contributed to our understanding of how collaborative engagement can encouraged within undergraduate biology classrooms and the types of discourse that may be most effective at promoting group learning. Implications from the first study are that many of the factors impacting human cooperation are reflected in the classroom. While this finding may seem obvious, the social environments of classrooms are very different than those that naturally occur in other contexts. Students are stratified into restricted classroom based social environments where they have limited ability to shape whether or not they choose to engage with peers and who they engage with. Also in the classroom the expectations of such engagements are not the same as those they would natural engage in during everyday life. The purpose of cooperation in the classroom is not about a mutually shared and internalized purposes (i.e. you cooperate because you chose to). Instead cooperation in the classroom is dictated from an external source (e.g. teacher) and thus not primarily emerging from an internal motivation. As such, the ability of instruction to support internal motivations to cooperate over instructor driven pressure may be paramount in promoting rich collaborative
engagement. Benefit from peers (either perception of or through the interdependency intervention) was the immediate predictor of either student dispositions towards engaging with peers or observed cooperation throughout this dissertation. The idea that a benefit to cost calculus (whether conscious or not) is important to human social behavior is not new (Ruff & Fehr, 2014). Yet its use in understanding classroom dynamics, as well as in education in general, is not common despite its potential as a potent lens for understanding student engagement.

The potential usefulness of supporting a greater likelihood of benefit from peers can be seen in the results of the second and third studies of the dissertation. Positive shifts in not only the amounts of whole-classroom and individual student engagement with peers, also significant shifts in the ways in which students were choosing to engage (via discourse) were seen in response to the interdependency intervention. Thus it appears that supporting benefit may not only increase student willingness to engage, but also the ways in which collaborative engagement occurs. This shift was particular important in its ability to promote student justification of their ideas, via explanation, within conversation. This was not only linked to learning in the third study, but also aligns with work from a myriad of other educational contexts (Asterhan & Schwartz, 2009; Chiu, 2008; Hausmann, Chi, & Roy, 2004; Webb, 1991) and is additionally an important science communication skill.

If there is an overall theme to be taken from the work contained in this dissertation, it is that the social environment in which learning occurs is not tangential to instruction or curriculum, but an active participant in generating student experience and learning in the classroom. Experienced educators would most likely agree that some classes are more engaged than others. Yet understanding the sources and potential interventions that can simultaneously support less engaged classes and push for even greater collaborative engagement in already
engaged classes is critical. For this to occur, requires further commitment (especially at the undergraduate level) by scholars to exploring aspects to classroom social environments which mediate the relationship between activities and student learning.

The importance of this relationship has only risen in importance over the last decade. The movement to advance undergraduate biology education, as was discussed in the introduction to the dissertation, has resulted in more than just an increased focus on practices of science. Perhaps the most substantial shift has been in encouraging educators to move away from lecture style instruction and towards “active learning” approaches (Freeman et al., 2014). While the specifics of what constitutes active learning remains opaque and lacks explicit characterization, many of the instructional strategies associated with more active approaches (e.g. think-pair-shares, problem based learning, inquiry based learning, and course based undergraduate research experiences to just name a few) explicitly rely on student cooperation for their effectiveness. Thus the results of this dissertation not only speak to the effectiveness of one instructional approach, but also have implications for how students engagement with, and learn from, some of the most widely used and cutting edge instructional techniques in undergraduate biology education. Specifically, both the amount of student cooperation, and the ways in which students choose to cooperate through discourse, require explicit attention by instructors if optimal learning is to occur. What optimal results look like is unknown at this time and most likely will vary based on a multitude of factors. Yet not taking into account the social environment when the type of learning expected is fundamentally social in nature begs reconsideration. Thus making investigations into how social environments mediate the effectiveness of active learning approaches excellent terrain for future work.
REFERENCES


APPENDIX 1

Supplemental materials for:

PROMOTING COLLABORATIVE CLASSROOMS: THE IMPACTS OF INTERDEPENDENT COOPERATIVE LEARNING ON UNDERGRADUATE INTERACTIONS AND ACHIEVEMENT

Contents

Pages 145-153: Weekly question sets
Pages 154-156: Cooperative Classroom Environment Measure (CCEM)
Pages 157-158: Supplemental Table #1 Psychometric properties of the CCEM (Pre and Post)
Pages 159-161: Cooperative Classroom Observation Protocol (CCOP) with explanation form
Page 162: Supplemental Table #2 Psychometric properties of the CCOP
Pages 163-168: Weekly quizzes
Page 169: Supplemental Table #3 Psychometric properties of weekly quizzes
Page 170: Supplemental Table #4 Model selection using class level factors to predict late prosocial behavior
Page 171: Supplemental Table #5 Model selection using class level factors to predict late discussion practices

Student Group work

148
Week of February 6, 2017

The following questions are related to the paper from Liu et al. “Bacteriophages of wastewater foaming-associated filamentous Gordonia reduce host levels in raw activated sludge” that you read for this week. You will have a quiz next week, (Feb 13) related to this material.

**Question 1:** What is the next set of experiments that the authors could/should do related to this project?

**Question 2:** Here are 3 sentences from the manuscript. If you were writing a manuscript, which section (introduction, Materials and Methods, Results, Discussion) would you place each statement in? Justify your answer.

A. “All isolates developed colonies with irregular margins and appeared white to beige at the beginning of incubation. Strain G1 produced a pink pigment, and G5 and G11 produced yellow and orange/red pigment, respectively, over prolonged incubation.”

B. “Phages against filamentous bacteria, especially different species of mycolata, have been isolated and their therapeutic applications documented in several studies. Mycobacteria species in particular have been the subject of focused efforts. Multiple phages against Rhodococcus equi were characterized and demonstrated to be capable of reducing R. equi load in a soil matrix.”

C. “… application of these phages resulted in repeatable, significant suppression of Gordonia levels in activated sludge conditions. This is surprising, given the considerable diversity and species richness observed in the micro- and macro organism community of the activated sludge, with different organismal groups responsible for complex functions including floc-forming, phosphorus removal, nitrite oxidation, and denitrification.”

**Question 3:** Consider the following questions and answer them using the knowledge you have gained in the SEA-PHAGES lab.

A. The host range determination experiment most closely resembles which experiment that you will carry out in Biology 107?

B. Why did the authors centrifuge the wastewater sample as their first step?

C. Why did the authors use 0.22 μM filters in their purification steps?
**Student Group Work**

Week of February 13

This week’s group work is related to Bacteria and Bacteriophages. You will take a quiz on this material during the week of Feb 20.

1. A. What kind of experiment is this?  
   B. Discuss the similarities and differences between Points a, c, and d on the plate above.  
   What biological organism(s) is present in each location, and where did they come from in the experiment?

2. A. Describe the features or attributes that make an organism “living”.  
   B. Using what you have learned so far this semester, would you consider bacteria to be living?  
   C. Using what you have learned so far this semester, would you consider viruses, and in particular bacteriophage, to be living?

3. A pinworm is a parasite that infects humans and lives and grows within our digestive tract.  
   A. In the pinworm-human interaction, which organism is the host and which is the parasite?  
   B. What does the host provide to the parasite?  
   C. How is the human-pinworm example related to bacteriophage and M. smegmatis? Describe the similarities and differences.
This week’s group work is related to the Logic of Experiments.

1. A student is trying to decide whether or not to use a plaque assay or a spot test during their phage isolation. Which one(s) should they use under the following scenarios? Why?

A. The student has just completed an Enrichment

B. The student has just completed Direct plating and has a putative phage

2. A student is performing a spot test for a potential phage that they have isolated. On their plate, they spot a phage sample and there are two additional solutions that they spot. Which one is a positive control, and which one is a negative control?

A. The student spots 5 uL of their phage buffer

B. The student spots 5 uL of phage buffer with mycobacteriophage D29, a previously discovered phage.

3. A student is isolating a phage and on their first attempt they get the plate labeled “i” below. The student makes a single change in their experimental approach and they get the plate labeled “ii”.

A. What do you think the single change in the experimental procedure was?

B. Why would a student want to use plate ii instead of plate i if they were to move forward with phage purification?
This week’s group work is related to Experimental Troubleshooting. You will take a quiz on this material during the week of March 6.

1. A student is purifying their phage and they get a plate like that shown on the right. They do two follow up plates. For the first they pick a large plaque and get small and large plaques on their follow up plate. For the second they pick a small plaque and get only small plaques on their follow up plate. Explain these observations.

2. A student performs a spot test of their putative phage and sees a plaque. In the negative control they see a plaque, and in their positive control they see no plaques. What conclusions could be reached about this experiment?

3. A student is performing serial dilutions during phage purification. They take their original phage sample ($10^0$) and want to make 100 uL of a 10-fold dilution ($10^{-1}$) solution. They have some tubes with 90 uL of phage buffer and others with 100 uL of phage buffer. Which tube do they choose, and why?
Student Group Work

Week of March 6

This week’s group work is related to using Phage as tools in biology. You will take a quiz on this material during the week of March 20 (AFTER SPRING BREAK!).

1. One application of phages is to quickly detect the presence of food-borne pathogenic bacteria in food. As an example, a microbiologist working for Chobani takes two batches of yogurt off of the production line. She suspects that one yogurt sample is contaminated with *Listeria*, a common food-borne pathogen, while the other one may or may not be contaminated.

   A. The scientist takes an extract from the contaminated yogurt sample and adds 1000 *Listeria*-specific virus particles. After shaking and incubation for a day, what will happen to the number of phage particles in the sample? Why? Is this experiment selective for *Listeria* or could other bacteria lead to false positive results?

   B. Based on your answer to Part A., how could the scientist tell if the second yogurt sample was contaminated with either *Listeria* or with *E. coli*, another potentially pathogenic bacterial species?

2. Phage therapy is an alternative to antibiotic treatment for bacterial infections. In phage therapy, phages specific to pathogenic bacteria are delivered to the site of an infection (e.g. on wound dressings, by oral ingestion, through an IV).

   A. One side effect of antibiotic treatment is that these chemicals kill many beneficial bacteria in the human gut. Why might phage therapy, for example to combat a *Listeria* infection in the gut, be superior?

   B. One problem with antibiotics is that they are typically unstable and quickly degrade in the body resulting in the need for frequent, high doses during treatment. What advantage would phage therapy have over antibiotic treatments in this respect?

3. *Leuconostoc* is a type of bacteria that is used widely in food fermentation, including during the production of wine. Many reports have surfaced of phages negatively influencing wine making by killing *Leuconostoc* bacteria.

   A. Discuss with your partners if you think either lytic, lysogenic, or both types of phages would have a major, negative impact on wine making.

   B. How could wine makers reduce their chances of losing *Leuconostoc* bacteria during wine making?
This week’s group work is related to the Replication of Experiments. The first two questions are related to the notebook entry below.

“Feb 1, 2015. Title: Enrichment of Environmental Samples. Procedure: A soil sample was obtained. After shaking and incubation, the culture was transferred into a 3 ml syringe filer unit using aseptically using a transfer pipet. Using a micropipettor dispense 50 uL of the undiluluted enrichment sample into a culture tube and mix well. After this allow the tube to sit at room temperature for 5 minutes and allow the phage to infect bacteria. Next, draw a grid on the bottom of an agar plate and label blocks of the grid by the positive control, negative control, and each of the dilutions of the enrichment sample. Obtain heated top agar, and after adding bacteria evenly spread the agar of the agar plate and allow the plates to sit for ten minutes. After cooling and solidification of the agar plates. Transfer 5 uL of the negative control using a micropipettor to the negative labeled block of the surface of your plate. Follow by transferring 5 uL of positive control and enrichment culture dilution to the positive labeled blocks on your agar plate and incubate after allowing the agar to solidify. Check these plaques after 24 hours.

Feb 7, 2015. Plaques obtained on positive control and environmental sample, no plaques for negative control.”

1. You are looking for a phage that infects Mycobacterium to use in phage therapy. The passage above was a student group’s notebook entry related to the initial phase of phage isolation. Is there additional data you would need to request before you would consider using that phage in your experiments? If yes, what data would you need to collect?

2. You decide to try and reproduce the laboratory experiment given above. You and your team find different results where your negative and positive controls appear to work correctly, but you get no phage from the environmental sample. What conclusions can be drawn from these two findings?

3. Why is it important that experimental observations be written down accurately and completely?
Student Group Work

Week of March 27

This week’s group work is related to Serial Dilutions.

1. You are counting plaques on your plaque assay plates made from serial dilutions of your high titer lysate. Your $10^{-5}$ plate has 615 plaques although some are butting up against each other so it is a judgment call. Your $10^{-6}$ plate has 42 plaques, and your $10^{-7}$ plate has only 1 plaque. Which plate would probably yield the most accurate titer calculation of your phage and why is it more trustworthy than the others?

2. You are given a phage lysate and a culture of bacterial cells. You are asked to determine the titer (# of phage/mL). You make three serial dilutions of 100-fold and a final 10-fold dilution of the sample. After infecting the host with 0.5 mL of the last dilution and plating with top agar, the lawn of bacteria generate 40 plaques. What was the titer of the original phage sample?

3. You and your partner are isolating a new phage and have produced both a High titer lysate of $\sim 10^{11}$ phage/mL and a Medium titer lysate of $\sim 10^{5}$ phage/mL. You have your tubes with the HTL and MTL sitting on your bench when your partner accidentally squirts ethanol on the two tubes, erasing all of the non-permanent sharpie markings on them. Both tubes have very similar volumes and look identical in nearly every way. What is a quick experiment you could perform that would allow you to assess which tube held the MTL and which one held the HTL without using up too much of your samples?
Student Group Work

Week of April 3-Plagiarism

This week’s group work is related to Plagiarism.

1) A student likes the way a statement is written in a manuscript, and wants to use that statement in their own lab report introduction. What precautions should they take to be sure they are not going to plagiarize the material?

2) A student emails their lab report to a friend for ideas about how to write the discussion section of their lab report. The student’s friend copies some of the student’s discussion word for word, and when the assignments are turned in, the instructor detects the identical sentences in the SafeAssign plagiarism software used in Biology 107. Who is at fault for this act of plagiarism, and should the punishments be the same for both the student and their friend?

3) What are the reasons that students plagiarize materials? What precautions can they take to minimize the risk?
Student Group Work

Week of April 10

This week’s group work is related to DNA Preparation. You will take a quiz on this material during the week of April 17.

1) The phage capsid protects phage DNA from environmental factors. During the process of phage DNA preparation, how can you take advantage of the capsid to assure acquisition of phage DNA without contaminating DNA from the bacterial host in the lysate?

2) Guanidinium thiocyanate denatures proteins, making it hazardous to users and is used in preparation of phage DNA. What role(s) does the process of protein denaturation play in the DNA purification process, and what precautions can be used by students to minimize their exposure to this chemical?

3) Some students get poor DNA recovery during the phage DNA preparation, with little DNA to show for their efforts. Why might some student pairs have low DNA yields while others have strong yields?
<table>
<thead>
<tr>
<th>#</th>
<th>Question</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neither Agree Nor Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Helping a classmate when they need help will ensure they help me when I need help.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>I am willing to help classmates outside of class if they need it.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>I know my classmates from outside of class.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>Class is more enjoyable when I work with other students.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>The more classmates participate in class discussions, the more I understand.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>My reputation in class is something that I value.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>I feel that I need to cooperate with my classmates.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>If I help a classmate with a question they will help me with other questions later.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>Friendships I have in this class also exist outside of class.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>I would rather help a classmate when I finish my work than sit around and wait.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>I learn best when working with classmates.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>12</td>
<td>Classmates’ ideas positively increase my learning experience.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>13</td>
<td>Classmates I help tend to help me back.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>My classmates expect me to cooperate with them.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>It is expected that I will work well with my classmates.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I have friends in class that I spent time with outside of class.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I put more energy into working cooperatively than my classmates.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I spend time helping my classmates during class.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I would rather work alone than with a partner.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I spend a greater amount of time helping classmates than I get helped.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I receive better grades in classes where I work with other students.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>When classmates share their ideas in class it helps me learn.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I am tend to spend more energy thinking of good ideas than do my classmates.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>If I help a classmate with their homework they will help me with mine later.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I prefer to take classes where students work together to solve problems.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>My classmates spend less time helping me than I help them.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I want to have a good reputation in my classes.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I expend more personal resources during cooperative exchanges than my classmates do.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Question</td>
<td>Options</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>It is assumed that I will be cooperative towards others in the classroom.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>The amount I understand is increased by classmate ideas.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>I care what my classmates think of me.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>I help other classmates during class when they need help</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>I dedicate energy to making sure my classmates receive the help they need.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>What gender do you identify with? / What is your gender?</td>
<td>Male(1) Female(2) Other(3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>What is your race?</td>
<td>White/Caucasian(1) Latino/a(2) Asian/Pacific Islander(3) African/African American(4) Other/ Multiracial (5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>Approximately how many undergraduate science courses have you completed prior to this semester?</td>
<td>0 (1) 1-2 (2) 3-5 (3) 6-10 (4) More than 10 (5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>Are you currently interested in pursuing a profession in a science related field?</td>
<td>Yes (1) No (2) Undecided (3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Supplemental Table 1
Confirmatory factor analysis results for Cooperative Classroom Environments Measure (CCEM) Pre-test and Post-test

<table>
<thead>
<tr>
<th>Subscale (Cronbach’s alpha pre/post)</th>
<th>Items</th>
<th>Pre-test Subscale Loadings</th>
<th>Post-test Subscale Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reciprocity</td>
<td>Helping a classmate when they need help will ensure they help me when I need help.</td>
<td>.78*</td>
<td>.75*</td>
</tr>
<tr>
<td></td>
<td>If I help a classmate with a question they will help me with other questions later.</td>
<td>.84*</td>
<td>.80*</td>
</tr>
<tr>
<td></td>
<td>Classmates I help tend to help me back.</td>
<td>.46*</td>
<td>.58*</td>
</tr>
<tr>
<td></td>
<td>If I help a classmate with their homework they will help me with mine later.</td>
<td>.68*</td>
<td>.73*</td>
</tr>
<tr>
<td>Willingness to help peers</td>
<td>I am willing to help classmates outside of class if they need it.</td>
<td>.58*</td>
<td>.51*</td>
</tr>
<tr>
<td></td>
<td>I would rather help a classmate when I finish my work than sit around and wait.</td>
<td>.53*</td>
<td>.53*</td>
</tr>
<tr>
<td></td>
<td>I spend time helping my classmates during class.</td>
<td>.61*</td>
<td>.55*</td>
</tr>
<tr>
<td></td>
<td>I help other classmates during class when they need help.</td>
<td>.60*</td>
<td>.64*</td>
</tr>
<tr>
<td></td>
<td>I dedicate energy to making sure my classmates receive the help they need.</td>
<td>.69*</td>
<td>.62*</td>
</tr>
<tr>
<td>Friendship presence</td>
<td>I know my classmates from outside of class.</td>
<td>.78*</td>
<td>.74*</td>
</tr>
<tr>
<td></td>
<td>Friendships I have in this class also exist outside of class.</td>
<td>.69*</td>
<td>.85*</td>
</tr>
<tr>
<td></td>
<td>I have friends in class that I spent time with outside of class.</td>
<td>.86*</td>
<td>.79*</td>
</tr>
<tr>
<td>Preference for cooperation</td>
<td>Class is more enjoyable when I work with other students.</td>
<td>.74*</td>
<td>.72*</td>
</tr>
<tr>
<td></td>
<td>I learn best when working with classmates.</td>
<td>.84*</td>
<td>.82*</td>
</tr>
<tr>
<td></td>
<td>I would rather work alone than with a partner. (inverse)</td>
<td>.66*</td>
<td>.65*</td>
</tr>
<tr>
<td></td>
<td>I receive better grades in classes where I work with other students.</td>
<td>.75*</td>
<td>.65*</td>
</tr>
<tr>
<td></td>
<td>I prefer to take classes where students work together to solve problems.</td>
<td>.77*</td>
<td>.77*</td>
</tr>
<tr>
<td>Benefit from classmate ideas</td>
<td>The more classmates participate in class discussions, the more I understand.</td>
<td>.64*</td>
<td>.69*</td>
</tr>
<tr>
<td></td>
<td>Classmates’ ideas positively increase my learning experience.</td>
<td>.72*</td>
<td>.78*</td>
</tr>
<tr>
<td></td>
<td>When classmates share their ideas in class it helps me learn.</td>
<td>.84*</td>
<td>.80*</td>
</tr>
<tr>
<td></td>
<td>The amount I understand is increased by classmate ideas.</td>
<td>.72*</td>
<td>.78*</td>
</tr>
<tr>
<td>Factor</td>
<td>Item</td>
<td>Alpha1</td>
<td>Alpha2</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>Reputational concern</td>
<td>My reputation in class is something that I value.</td>
<td>.79*</td>
<td>.84*</td>
</tr>
<tr>
<td></td>
<td>I want to have a good reputation in my classes.</td>
<td>.82*</td>
<td>.79*</td>
</tr>
<tr>
<td></td>
<td>I care what my classmates think of me.</td>
<td>.62*</td>
<td>.72*</td>
</tr>
<tr>
<td>Cooperative Norms</td>
<td>I feel that I need to cooperate with my classmates.</td>
<td>.70*</td>
<td>.72*</td>
</tr>
<tr>
<td></td>
<td>My classmates expect me to cooperate with them.</td>
<td>.62*</td>
<td>.72*</td>
</tr>
<tr>
<td></td>
<td>It is expected that I will work well with my classmates.</td>
<td>.65*</td>
<td>.66*</td>
</tr>
<tr>
<td></td>
<td>It is assumed that I will be cooperative towards others in the classroom.</td>
<td>.59*</td>
<td>.59*</td>
</tr>
<tr>
<td>Relative investment in cooperation</td>
<td>I put more energy into working cooperatively than my classmates.</td>
<td>.34*</td>
<td>.41*</td>
</tr>
<tr>
<td></td>
<td>I spend a greater amount of time helping classmates than I get helped.</td>
<td>.74*</td>
<td>.64*</td>
</tr>
<tr>
<td></td>
<td>I am tend to spend more energy thinking of good ideas than do my classmates.</td>
<td>.56*</td>
<td>.66*</td>
</tr>
<tr>
<td></td>
<td>My classmates spend less time helping me than I help them.</td>
<td>.73*</td>
<td>.61*</td>
</tr>
<tr>
<td></td>
<td>I expend more personal resources during cooperative exchanges than my classmates do.</td>
<td>.50*</td>
<td>.62*</td>
</tr>
</tbody>
</table>

Note that the factor loadings are correlations between student answers for an item and the subscale factor (e.g. reciprocity). A factor loading is can be considered adequate if >.30. *p < .05
<table>
<thead>
<tr>
<th>#</th>
<th><strong>Indicators of Prosocial Behavior</strong></th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neither Agree Nor Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Students do not behave in ways which are distracting to peers.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>Students display behaviors indicative of listening to one another during discussion.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>Students voluntarily offer ideas in the classroom without being prompted.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>Students show support for their peer’s ideas during discussion.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>Students appear to enjoy working with one another.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>#</th>
<th><strong>Discussion Practices</strong></th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neither Agree Nor Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Students reference what their peers are saying when presenting their own ideas.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>Students build on their classmate’s ideas during discussion.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>Students provide evidence to support their ideas.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>Students go into depth with ideas when discussing.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
Observation Protocol Item Descriptors

Indicators of Prosocial behavior (items in bold, descriptors underneath)

Students do not behave in ways which are distracting to peers.

- Students are not doing the following:
  - doing other work
  - on phone (including texting).
  - talking about non-content topics

Students display behaviors indicative of listening to one another during discussion.

- Students are generally doing the following:
  - Have their heads up when during discussion (unless writing)
  - Turn to face speakers.

Students voluntarily offer ideas in the classroom without being prompted.

- Prompts can be from either the instructor or fellow students (specifically in small groups or pairs)
  - Example: Once the instructor asks a question multiple students answer without having to ask for answers in between each question.
  - Example 2: Both students (pairs) or all students (groups) discuss without some students never contributing (social loafing).

Students show support for their peers ideas during discussion.

- Adds nothing to the science content specifically but shows peer support for either each other towards task.
  Includes:
  - Agreeing head nods
  - Verbal support (including praise)
    - Examples: “good idea” “I agree”
  - Helping out another student that has a question or can’t answer a question
    - Example 1: A student asks a question and another student answers it.
    - Example 2: A student can’t come up with an question asked by another student or the instructor and another student helps them by answering it or contributing to an answer.

Students appear to enjoy working with one another.

- This item is focused on behavioral expressions of student enjoyment. Includes:
  - Smiles
  - Laughter
  - General positive feel and energy in student interactions
Discussion Practices

Students reference what their peers are currently saying when presenting their own ideas.

- This item looks at continuity between contributors in the conversation. (explicitly)
  o Example: Instructor asks “What do you know about viruses?” John says “I know that viruses are small..” and Jane adds “I agree with John and also _______”
  o NON-Example: Instructor asks “What do you know about viruses?” John says “I know that viruses are small..” and Jane says “They aren’t alive”

Students build on their classmate’s ideas during discussion.

- This item also looks at continuity between contributors in the conversation.(implicitly)
  o Example: Instructor asks “What do you know about viruses?” John says “I know that viruses are small..” and Jane adds “Yeah they are small and also _______”
  o NON-Example: Instructor asks “What do you know about viruses?” John says “I know that viruses are small..” and Jane says “They aren’t alive”

Students provide evidence to support their ideas.

- This item looks at student sources of ideas with more evidence being present being an indication of more effective discussion. All types are equal in this item so any type can support agreement.
  o Sources of evidence can include textbook, instructor statement, lecture/lab material, logic, internet, experience, logic etc.
    ▪ Example for source evidence: Instructor asks “What do you know about viruses?” John says “I read in the textbook/notes/ Dr. Davis said in lecture that viruses are small”
    ▪ Example for experience evidence: Instructor asks, “What is the next step we should take in X situation?” Tammy replies, “Our group had that happen last week and we did ____”
    ▪ Example for logic evidence: Instructor asks, “What is the next step we should take in X situation?” Zeke replies, “Well we just finished doing X so we need to complete protocol Y to see the results”

Students go into depth with ideas when discussing.

- This item also looks at whether students will remain on the surface level of an idea or actually unpack the details related to it. If you have to prompt students to do this than it does not count.
  o Example: Instructor asks “What do you know about viruses?” James says “I know that viruses are not alive…” Lisa responds “How do we determine that?” Ashley responds, “I think they use ______”
  o NON-Example: Instructor asks “What do you know about viruses?” James says “I know that viruses are not alive..” and Lisa says “Yeah and they are small”
<table>
<thead>
<tr>
<th>Subscale</th>
<th>Item</th>
<th>Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Indicators of Prosocial Behavior</strong></td>
<td><strong>Students do not behave in ways which are distracting to peers.</strong></td>
<td>.29*</td>
</tr>
<tr>
<td>(α = .70)</td>
<td><strong>Students display behaviors indicative of listening to one another during discussion.</strong></td>
<td>.61*</td>
</tr>
<tr>
<td></td>
<td><strong>Students voluntarily offer ideas in the classroom without being prompted.</strong></td>
<td>.57*</td>
</tr>
<tr>
<td></td>
<td><strong>Students show support for their peer’s ideas during discussion.</strong></td>
<td>.72*</td>
</tr>
<tr>
<td></td>
<td><strong>Students appear to enjoy working with one another.</strong></td>
<td>.55*</td>
</tr>
<tr>
<td><strong>Discussion Practices</strong></td>
<td><strong>Students reference what their peers are saying when presenting their own ideas.</strong></td>
<td>.50*</td>
</tr>
<tr>
<td>(α = .74)</td>
<td><strong>Students build on their classmate’s ideas during discussion.</strong></td>
<td>.67*</td>
</tr>
<tr>
<td></td>
<td><strong>Students provide evidence to support their ideas.</strong></td>
<td>.68*</td>
</tr>
<tr>
<td></td>
<td><strong>Students go into depth with ideas when discussing.</strong></td>
<td>.70*</td>
</tr>
</tbody>
</table>

Note that the factor loadings are correlations between student answers for an item and the subscale factor (e.g. Discussion Practices). A factor loading is can be considered adequate if > .30.

\*p < .05
Biology 107 Laboratory

Spring 2017

Quiz 1 – Week of February 13, 2017

2 Points total; 0.5 points per question

1. Below are two statements taken from a student’s laboratory report in Biology 107. In which section of the lab report should the student place each statement (Introduction, Materials and Methods, Results, Discussion)?

A. “Using the plate from the spot assay procedure, each plaque was labeled and circled. 90 uL of SM buffer was added to a microcentrifuge tube. Using a pipet tip to poke the plaque, the phage was transferred to the SM buffer microcentrifuge tube.”

B. “The calculated titer of $1.04 \times 10^9$ (pfu/mL) was slightly lower than the required $1.0 \times 10^{10}$ minimum to move on to isolating and purifying the genomic DNA of the phage. This low concentration may have been caused by the concentration of purified phage being transferred for the Titer Assay being too low as well.”

2. A common experimental approach in phage hunting is to centrifuge an environmental sample taken from, e.g. wastewater or a soil sample incubated in buffer. Why would a scientist centrifuge their sample before beginning a bacterial infection assay?

3. A scientist is working with a mixture of phage and bacteria. They grab a 5 um filter off the shelf and use it to filter a mixture of bacteria and phage. What will be present in the filtrate, and what will be trapped in the filter?
Biology 107 Laboratory

Spring 2017

Quiz 2 – Week of February 20, 2017

2 Points total; 0.5 points per question

1. Below are two statements. For each, fill in the blank with either bacteria, viruses, or both bacteria and viruses.

A. ____________________ are able to generate their own energy using only external sources such as sunlight or glucose. Therefore they are considered living.

B. ____________________ are able to infect other organisms. This does not tell a scientist if they are living or non-living.

2. If a student were to grow organisms from Points a and d on the experimental outcome shown above, what would they obtain?

<table>
<thead>
<tr>
<th>Point</th>
<th>Bacteria (Y/N)?</th>
<th>Phage (Y/N)?</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Quiz 3 – Week of March 6, 2017

2 Points total; 0.5 points per question

1. A student is performing a spot test and does both a negative and a positive control. For each situation below, indicate if the student should move forward with their experiment or not (Yes/No).

A. The student sees a plaque on their positive control and no spot on the negative control.

B. The student sees a spot on the negative control, but no spot on the positive control.

2. To make 100 uL of a $10^{-2}$ phage sample, a student should use ___________ uL of phage buffer and ___________ uL of a $10^{-1}$ phage stock.

3. A student is purifying their phage and they get a plate like that shown on the right. They pick a small, isolated plaque and get both small and large plaques on their follow up streak plate. After 3 more plates, they still get a mixture of large and small plaques on their streak plates.

At this point, which of these two scenarios is more probable? Why?

i. They have two different phages and they cannot separate them

ii. They have one phage that shows 2 different plaque morphologies that depend on the local environmental conditions in the top agar
Biology 107 Laboratory

Spring 2017

Quiz 4 – Week of March 20, 2017

2 Points total; 1 point each question

1. A food scientist is working for a dairy farm and she suspects that a dangerous strain of E. coli may have contaminated a batch of milk. Describe an experiment that would allow the scientist to quickly and selectively screen for the presence of E. coli.

2. Phage therapy has been shown to have advantages over traditional chemical antibiotics.

   A. What is one reason that salmonella phages be preferred over chemical antibiotics to treat a salmonella infection in the human gut?

   B. How many doses of salmonella phages would a doctor need to supply, in theory, if a patient came in with a salmonella infection? Compare this to the standard, 5-7 day course of antibiotics that are normally used.
Biology 107 Laboratory

Spring 2017

Quiz 5 – Week of March 27, 2017
Material related to Replication of Experiments

2 Points total; 1 point each question

1. You want to see which, if any, phages in your lab section will infect *Mycobacterium tuberculosis* as part of a phage therapy project. Your lab colleague hands you their notebook related to the phage that they have isolated and as you review the days reporting the isolation of their phage, you find the following information clearly noted:

- The GPS coordinates of two soil samples
- Experimental details of a direct enrichment experiment, including the bacterial species used
- Pictures of an enrichment plate indicating plaques from the positive control and several plaques in the experimental area of a plate.

If you needed to, could you reproduce this experiment? If not, what information is missing in the experimental details?

2. Your work this semester is part of a national research project related to bacteriophage discovery and characterization. Name two types of individuals who would be interested in your results and might have reason to read your lab notebook and who would want to be able to reproduce your experimental results
Quiz 6 – Week of April 17, 2017
Material related to Phage DNA Preparation

2 Points total; 0.5 points each question part

1. You have a liquid sample that contains intact phage particles and ruptured bacterial cells.
   a. Where would you expect to find DNA in this sample?
   b. How can you treat this sample so that at the end of your work you will maximize the yield of phage genome and minimize the yield of potential interfering nucleic acids?

2. Two student groups, labeled Group 1 and Group 2, performed the gDNA isolation protocol incorrectly. Group 3 performed the gDNA isolation protocol correctly.
   a. Group 1 treated the HTL lysate with resin before they added nuclease but followed all other steps. What would be the expected DNA concentration and purity for Group 1 as compared to Group 3?
   b. Group 2 forgot to add nuclease but otherwise completed followed all other steps. What would be the expected DNA concentration and purity for Group 2 as compared to Group 3?
Supplemental Table 4

*Confirmatory factor analysis model assessing the extent to which the weekly quizzes were assessing a common factor (hypothesized to be science content knowledge).*

<table>
<thead>
<tr>
<th>Quiz Number</th>
<th>Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>.51*</td>
</tr>
<tr>
<td>Two</td>
<td>.45*</td>
</tr>
<tr>
<td>Three</td>
<td>.39*</td>
</tr>
<tr>
<td>Four</td>
<td>.50*</td>
</tr>
<tr>
<td>Five</td>
<td>.09</td>
</tr>
<tr>
<td>Six</td>
<td>.32*</td>
</tr>
</tbody>
</table>

Note that quiz number five did not share significant common variance with other weekly quizzes and thus was not included in the analyses. Factor loadings are correlations between student answers for an item and the subscale factor (e.g. Discussion Practices). A factor loading is considered adequate if >.30.

*p < .05*
### Supplemental Table 4
Model selection using class level factors to predict late prosocial behavior

<table>
<thead>
<tr>
<th>Initial model and terms dropped</th>
<th>Changes in Model Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcome</strong>: late prosocial behavior</td>
<td>df</td>
</tr>
<tr>
<td><strong>Initial Model</strong>: rater level + initial prosocial behavior + course achievement + initial science career motivation + percentage female + percentage interest in a science career + percentage with low science experience + initial reciprocity + initial friendships + initial willingness to help peers + initial reputational concern + initial perception of cooperative norms + initial relative investment in cooperation + condition + change reciprocity + change friendships + change willingness to help peers + change reputational concern + change perception of cooperative norms + change in relative investment in cooperation + (change in reciprocity x change in relative investment in cooperation)</td>
<td>19.71</td>
</tr>
<tr>
<td>– change in relative investment in cooperation</td>
<td>1</td>
</tr>
<tr>
<td>– change in cooperative norms</td>
<td>1</td>
</tr>
<tr>
<td>– change reputational concern</td>
<td>1</td>
</tr>
<tr>
<td>– change willingness to help peers</td>
<td>1</td>
</tr>
<tr>
<td>– initial perception of cooperative norms</td>
<td>1</td>
</tr>
<tr>
<td>– initial reputational concern</td>
<td>1</td>
</tr>
<tr>
<td>– initial willingness to help peers</td>
<td>1</td>
</tr>
<tr>
<td>– percentage with low science experience</td>
<td>1</td>
</tr>
<tr>
<td>– percentage interest in a science career</td>
<td>1</td>
</tr>
<tr>
<td>– initial science career motivation</td>
<td>1</td>
</tr>
<tr>
<td>– rater level</td>
<td>1</td>
</tr>
</tbody>
</table>

**Final Model**: initial prosocial behavior + course achievement + percentage female + initial reciprocity + initial willingness to help peers + condition + change reciprocity + change friendships

---

*a Each factor subtracted from the initial model are shown in the far left column. The development of the model was cumulative moving downward from the initial model to the final model. Factors not meeting the removal criteria are not shown (and retained in the model). Removal criteria were met if 1) removal of the factor decreased model AIC by < 2 or 2) If ΔAIC was 2 > |x| > 0 then the factor was removed to retain the most parsimonious model. If removing a factor resulted in an increase of AIC > 2 this was retained in the model. 
Supplemental Table 5
Model selection using class level factors to predict late discussion practices

<table>
<thead>
<tr>
<th>Changes in Model Fit</th>
<th>df</th>
<th>Residual df</th>
<th>Residual Deviance</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Model: rater level + initial discussion practices + course achievement + initial science career motivation + percentage female + percentage interest in a science career + percentage with low science experience + initial reciprocity + initial friendships + initial willingness to help peers + initial reputational concern + initial perception of cooperative norms + initial relative investment in cooperation + condition + change reciprocity + change friendships + change willingness to help peers + change reputational concern + change perception of cooperative norms + change in relative investment in cooperation</td>
<td>15.36</td>
<td>74</td>
<td>18.18</td>
<td>170.54</td>
</tr>
<tr>
<td>– change in friendship</td>
<td>1</td>
<td>15.36</td>
<td>75</td>
<td>18.20</td>
</tr>
<tr>
<td>– change in cooperative norms</td>
<td>1</td>
<td>15.34</td>
<td>76</td>
<td>18.19</td>
</tr>
<tr>
<td>– initial willingness to help</td>
<td>1</td>
<td>15.30</td>
<td>77</td>
<td>18.23</td>
</tr>
<tr>
<td>– initial perception of cooperative norms</td>
<td>1</td>
<td>15.27</td>
<td>78</td>
<td>18.27</td>
</tr>
<tr>
<td>– initial investment in cooperation</td>
<td>1</td>
<td>15.25</td>
<td>79</td>
<td>18.28</td>
</tr>
<tr>
<td>– percentage female</td>
<td>1</td>
<td>14.96</td>
<td>80</td>
<td>18.58</td>
</tr>
<tr>
<td>– percentage interest in a science career</td>
<td>1</td>
<td>14.77</td>
<td>81</td>
<td>18.77</td>
</tr>
<tr>
<td>– percentage with low science experience</td>
<td>1</td>
<td>14.70</td>
<td>82</td>
<td>18.84</td>
</tr>
<tr>
<td>– initial science career motivation</td>
<td>1</td>
<td>14.64</td>
<td>83</td>
<td>18.90</td>
</tr>
<tr>
<td>– initial discussion practices</td>
<td>1</td>
<td>14.95</td>
<td>84</td>
<td>19.20</td>
</tr>
</tbody>
</table>

Final Model: rater level + course achievement + initial reciprocity + initial friendships + initial reputational concern + condition + change reciprocity + change willingness to help peers + change reputational concern + change in relative investment in cooperation

*a Each factor subtracted from the initial model are shown in the far left column. The development of the model was cumulative moving downward from the initial model to the final model. Factors not meeting the removal criteria are not shown (and retained in the model). Removal criteria were met if 1) removal of the factor decreased model AIC by < 2 or 2) If ΔAIC was 2 > |x| > 0 then the factor was removed to retain the most parsimonious model. If removing a factor resulted in an increase of AIC >2 this was retained in the model.
APPENDIX 2

Supplemental materials for:

DISCOURSE REMIXED: USING INTERDEPENDENCY TO SHIFT STUDENT LEARNING THROUGH TALK

Contents

Pages 173-174: outcome measure 1 (Quiz) from Premo, Cavagnetto, & Davis, 2018 (Sup. Fig. 1) and the modified version (Sup. Fig. 2) used in the present study

Pages 175-176: outcome measure 2 (Quiz) from Premo, Cavagnetto, & Davis, 2018(Sup. Fig. 3) and the modified version (Sup. Fig. 4) used in the present study

Page 177: Question set #1 (Sup. Fig. 5)

Page 178: Question set #2 (Sup. Fig. 6)

Pages 179-180: Supplemental Table 1: Sub-code categories and code descriptions
Biology 107 Laboratory

Spring 2017 (from Premo, Cavagnetto, & Davis, 2018)

Quiz—Week of February 20, 2017

2 Points total; 0.5 points per question

1. Below are two statements. For each, fill in the blank with either bacteria, viruses, or both bacteria and viruses.

A. ________________________ are able to generate their own energy using only external sources such as sunlight or glucose. Therefore they are considered living.

B. ________________________ are able to infect other organisms. This does not tell a scientist if they are living or non-living.

2. If a student were to grow organisms from Points a and d on the experimental outcome shown above, what would they obtain?

<table>
<thead>
<tr>
<th>Point</th>
<th>Bacteria (Y/N)?</th>
<th>Phage (Y/N)?</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Biology 107 Laboratory
Fall 2017 (modified from Premo, Cavagnetto, & Davis, 2018)

Quiz – Week of October 1, 2017

2 Points total

1. Describe the similarities and differences between bacteria and bacteriophages. Be sure to highlight whether each species is considered alive or not (and a rationale for why for each type of organism), and what features of their interaction in the environment are consistent with a parasitic relationship.

2. If a student were to grow organisms from Points b and d on the experimental outcome to the left, what would they obtain?

<table>
<thead>
<tr>
<th>Point</th>
<th>Bacteria (Y/N)?</th>
<th>Phage (Y/N)?</th>
</tr>
</thead>
<tbody>
<tr>
<td>b.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Biology 107 Laboratory

Spring 2017 (from Premo, Cavagnetto, & Davis, 2018)

Quiz – Week of March 20, 2017

2 Points total; 1 point each question

1. A food scientist is working for a dairy farm and she suspects that a dangerous strain of E. coli may have contaminated a batch of milk. Describe an experiment that would allow the scientist to quickly and selectively screen for the presence of E. coli.

2. Phage therapy has been shown to have advantages over traditional chemical antibiotics.

   A. What is one reason that salmonella phages be preferred over chemical antibiotics to treat a salmonella infection in the human gut?

   B. How many doses of salmonella phages would a doctor need to supply, in theory, if a patient came in with a salmonella infection? Compare this to the standard, 5-7 day course of antibiotics that are normally used.
Quiz – Week of October 22, 2017

2 Points total

1. A food scientist is working for a dairy farm and she suspects that a dangerous strain of E. coli may have contaminated a batch of milk. Describe an experiment that would allow the scientist to quickly and selectively screen for the presence of E. coli. (0.5 point)

2. A patient arrives in a medical clinic complaining of symptoms that are consistent with an infection by bacteria called Salmonella, but the exact bacterial species has not been determined yet. The medical staff has two options to treat the patient-Salmonella phages or broad spectrum antibiotics. What are the advantages and disadvantages of the two treatment options (antibiotics vs. phage therapy)?

(1 point)

3. Following up from the scenario outlined in Question 2, how many doses of salmonella phages would a doctor need to supply, in theory, if a patient came in with a salmonella infection? Compare this to the standard, 5-7 day, daily pill course of antibiotics that are normally used.

(0.5 point)
**Student Group Work**

Week of September 24th

This week’s group work is related to Bacteria and Bacteriophages. You will take a quiz on this material during the week of Oct. 1st.

1. A. What kind of experiment is this?
   
   B. Discuss the similarities and differences between Points a, c, and d on the plate above. What biological organism(s) is present in each location, and where did they come from in the experiment?

2. A. Describe the features or attributes that make an organism “living”.
   
   B. Using what you have learned so far this semester, would you consider bacteria to be living?
   
   C. Using what you have learned so far this semester, would you consider viruses, and in particular bacteriophage, to be living?

3. A pinworm is a parasite that infects humans and lives and grows within our digestive tract.
   
   A. In the pinworm-human interaction, which organism is the host and which is the parasite?
   
   B. What does the host provide to the parasite?
   
   C. How is the human-pinworm example related to bacteriophage and M. smegmatis? Describe the similarities and differences.
**Student Group Work** (Week of October 15th)

This week’s group work is related to using Phage as tools in biology. You will take a quiz on this material during the week of October 22nd.

1. One application of phages is to quickly detect the presence of food-borne pathogenic bacteria in food. As an example, a microbiologist working for Chobani takes two batches of yogurt off of the production line. She suspects that one yogurt sample is contaminated with *Listeria*, a common food-borne pathogen, while the other one may or may not be contaminated.

   A. The scientist takes an extract from the contaminated yogurt sample and adds 1000 *Listeria*-specific virus particles. After shaking and incubation for a day, what will happen to the number of phage particles in the sample? Why? Is this experiment selective for *Listeria* or could other bacteria lead to false positive results?

   B. Based on your answer to Part A., how could the scientist tell if the second yogurt sample was contaminated with either *Listeria* or with *E. coli*, another potentially pathogenic bacterial species?

2. Phage therapy is an alternative to antibiotic treatment for bacterial infections. In phage therapy, phages specific to pathogenic bacteria are delivered to the site of an infection (e.g. on wound dressings, by oral ingestion, through an IV).

   A. One side effect of antibiotic treatment is that these chemicals kill many beneficial bacteria in the human gut. Why might phage therapy, for example to combat a *Listeria* infection in the gut, be superior?

   B. One problem with antibiotics is that they are typically unstable and quickly degrade in the body resulting in the need for frequent, high doses during treatment. What advantage would phage therapy have over antibiotic treatments in this respect?

3. *Leuconostoc* is a type of bacteria that is used widely in food fermentation, including during the production of wine. Many reports have surfaced of phages negatively influencing wine making by killing *Leuconostoc* bacteria.

   A. Discuss with your partners if you think either lytic, lysogenic, or both types of phages would have a major, negative impact on wine making.

   B. How could wine makers reduce their chances of losing *Leuconostoc* bacteria during wine making?
### Supplemental Table 1
Sub-code categories and code descriptions

<table>
<thead>
<tr>
<th>Category</th>
<th>Sub-Codes</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>On Task</strong></td>
<td>Yes [1] and No [0]</td>
<td>The purpose of these three coding categories is to provide information on the target and the nature of the student participation. <strong>On/Off Task</strong> – separates talk turns that are in alignment with the learning activity from those that are not. On task talk turns include contributions, questions, logistic statements, and other statements that are oriented towards completing the task. Off task talk turns are those that have no relationship to completing the task or task material in any way. Common themes for off task talk turns include discussing grades/assignment/instruction in the lecture portion of the course or another course, social networking with peers, and discussions of external media (movies, TV shows, etc.). <strong>Note</strong> that analysis was restricted to the period of conversation in-between the first on task statement to the last in the group. Thus any talk turn coded as off task took place during group completion of the task.</td>
</tr>
<tr>
<td><strong>Content Present</strong></td>
<td>Yes [1] and No [0]</td>
<td><strong>Content Present</strong> – includes any statement demonstrating science knowledge or requires an understanding of a science principle, idea, or terminology. <strong>Note</strong> that the connection to content must be explicitly present during the talk turn.</td>
</tr>
<tr>
<td><strong>Input Type</strong></td>
<td>Statement [S] and Question [Q] and Question Statement [QS]</td>
<td><strong>Input Type</strong> – this separates talk turns that are put forward as a question by the speaker from those that are statements. <strong>Note</strong> that questions that either offer an idea or include a justification are both coded as a question even though functionally the individual may be offering the idea in the form of a question in order to save face in the instance of their contribution being wrong. (Note that Question X Content Present [1] = questions that functionally contribute ideas to the conversation in addition to being questions.) <strong>Note</strong> that QS is a subset of the Question code. A question statement is an utterance posed as a question but is functioning as a statement. Students will sometimes do this to soften an idea. I think this is worth coding simply because it may tell us something in a secondary analysis of group dynamics. A cue for this is if there is a question that includes a justification - then assume they are couching a statement in a question for a social reason. <strong>Note</strong>: there are instances in which in a single utterance an individual will ask a clear question and a statement that is distinct from the question.</td>
</tr>
</tbody>
</table>
| **Source of Contribution** | From Task Materials [T], Instructor [I], Repetition [R]. Unique/no source apparent [U.] | The purpose of this coding category is to provide information on the resources that students draw upon for content within the discussion. This can provide information on the impact and usefulness of the provided materials, their peer’s ideas, or instructor to the conversation at hand. **Note** that only talk turns that are On Task [1] and have content present in them [1] are coded here. This is because the function of the code category is to specifically examine sources of contributions, thus the talk turn must be on task and provide a content relevant contribution in order to move the conversation forward in completing the task. **Task Materials** – includes talk turns where a student is explicitly drawing on information provided by task materials. Materials can include written information. **Instructor** – includes talk turns that follow a student who has either just interacted with, or listened to, an instructor AND is using information from this to contribute to group conversation. **Repetition** – includes talk turns whose contributions are coming from a previous contribution of a peer. **Note** that these are not bound to just what was said in the previous talk turn, but span the entire conversation. **Unique/no external source apparent** – this code is designed to identify the contributions most likely to arise from a student’s prior experience with and knowledge of the science content needed to complete the task. This code is used when a contribution does not appear to come from the other codes (Task Materials, Instructor, or Peers). **Note** that often students will repeat part of what a peer has contributed before adding their own contribution to it. In this case
The unique contribution is more important than the repetition and thus the talk turn would be coded as a \[U\]. (Also note that the Idea Building code will allow us to track these specific instances where a student is repeating another student’s idea in order to add their own unique contribution to it.)

*Only apply code to when “content present” utterance.

<table>
<thead>
<tr>
<th>Content Accuracy</th>
<th>Accurate [A], Inaccurate [IA], combination of accurate and inaccurate [B], not enough information [N]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>The purpose of this coding category is to provide information on the accuracy of content utterances by students. This is particularly useful as we seek to determine how content understanding develops among the group over the discussion(s) and the dynamics of how students respond to contributions of varying alignment with ideas from the course. Note that we are specifically interested in the accuracy of contributions that are coming from the students themselves. Thus ideas arising from other sources (materials [T] or instructor [I]) are not coded here. Accurate – talk turns whose contributions align with current scientific understandings of the subject discussed. Inaccurate - talk turns whose contributions do not align with current scientific understandings of the subject discussed. Combination of accurate and inaccurate – talk turns whose contributions have both accurate and inaccurate aspects to them (at least one of each). Not enough information – talk turns whose contributions are either incomplete or too non-specific to judge their accuracy.</td>
</tr>
<tr>
<td></td>
<td>*Only apply code to when “content present” utterance.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Justification</th>
<th>Justification [1], No justification [0],</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>The purpose of this coding category is to provide information on the frequency with which students justify their contribution to the group. These Justifications are functionally an explanation to reinforce their contributions to the group. This can take a variety of forms including explanations for why a contribution is correct through citing a source, using a logical inference, or providing a content based rationale.</td>
</tr>
<tr>
<td></td>
<td>*Only apply code to when “content present” utterance.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Response to Content Contribution</th>
<th>Explicit Agreement [A], Disagreement [D], Both agree and disagree [B], Neither explicitly agree nor disagree [N]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>The purpose of this code is to provide information on the relative amounts of agreement and disagreement during the conversation and thus allow for an examination of the different impacts of both types of responses on the conversation overall. Through these the codes can provide us with information to determine task difficulty and also information on adherence to class agreed upon norms of interaction. Explicit agreement – includes talk turns where an individual is providing explicit support for the ideas in the previous talk turn. This often manifests through statements like “yeah” and “perfect”, making of affirmative noises, or repetitions of the previous contribution. Note that agreement is not always explicit in nature. Thus the idea building code will be used to code implicit agreement, as there must be some understanding and acceptance of an idea before a student will make use of it in order to further the development of ideas during the conversation. Disagreement – includes talk turns where an individual provides either explicit or implicit commentary on the inaccuracy of the prior contribution. This also can include more implicit disagreement through questioning the idea or assumptions underlying it. Both agree and disagree – this code is used when are both elements of both agreement and disagreement (above codes) in reference to the previous statement in the same subsequent talk turn. Neither explicitly agree nor disagree – this code is used when there is no conceptual connection between the talk turn to be coded and the previous talk turn. This is common code that typically occurs when a student offers an idea not aligned with the previous strand of conversation or returns to this conversation strand if the previous talk turn was unaligned itself. Note that students will often talk without reference to what was previously said, this practice often manifests as the neither agree nor disagree code.</td>
</tr>
<tr>
<td></td>
<td>*Only apply code to when “content present” utterance.</td>
</tr>
</tbody>
</table>