

11-17-2018

An analysis of classroom collusion using Latent Dirichlet Allocation

Charles B. Shrader

Iowa State University, cshrader@iastate.edu

Sue P. Ravenscroft

Iowa State University, sueraven@iastate.edu

Jeffrey Kaufmann

University of Minnesota

Follow this and additional works at: https://lib.dr.iastate.edu/management_conf

Part of the [Curriculum and Instruction Commons](#), [Higher Education Commons](#), [Management Sciences and Quantitative Methods Commons](#), [Statistical Models Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Shrader, Charles B.; Ravenscroft, Sue P.; and Kaufmann, Jeffrey, "An analysis of classroom collusion using Latent Dirichlet Allocation" (2018). *Management Conference Papers, Posters and Proceedings*. 1.
https://lib.dr.iastate.edu/management_conf/1

This Conference Proceeding is brought to you for free and open access by the Management at Iowa State University Digital Repository. It has been accepted for inclusion in Management Conference Papers, Posters and Proceedings by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

DECISION SCIENCES INSTITUTE

An analysis of classroom collusion using Latent Dirichlet Allocation

ABSTRACT

In this study, we use Latent Dirichlet Allocation to explore the reflections of students who faced a demanding classroom challenge, to which some responded by colluding. Our five-topic LDA solution describes the cheating event in terms of the nature of the course assignment itself, teams as a resource and support mechanism, the repercussions of cheating, and differences between majors or course tracks. The most relevant topics were the differences between the tracks and the repercussions of cheating. Teams and teammates also play a large role in the students' reflections. We conclude with the implications of these topics in future research.

KEYWORDS: Cheating, Collusion, Latent, Dirichlet, Allocation

INTRODUCTION

In late 2017, The *Washington Post* reported that over 80 marketing students were accused of cheating at a major Midwestern university (Bever, 2017). The accused students used a group-messaging application to share answers pertaining to classroom assignments. Unfortunately, cheating is not a rare occurrence. Business educators view cheating by college students as constituting a long-standing problem that is serious and widespread at both the graduate and undergraduate level (e.g., Cronan, Mullins & Douglas, 2018; McCabe & Trevino, 1996; Simha & Cullen, 2012; Simkin & McLeod, 2010; Wright, 2004; Klein, Levenburg, McKendall & Mothersell, 2007; McCabe, Butterfield & Trevino, 2006). Evidence suggests that the problem is worldwide (Ismail & Yussof, 2016). Cheating in business schools appears to occur no more or less than in other academic disciplines, however, business student attitudes toward what constitutes cheating has been found to be relatively more lax (Klein, Levenburg, McKendall, & Mothersell, 2007).

Cheating is most often defined as either purposefully obtaining access to another person's work without authorization, or as unpermitted collaboration on exams or collusion on assignments (Burrus, McGoldrick & Schuhmann, 2007; McCabe & Trevino, 1996; Sierra & Hyman, 2008) with intent to gain advantage over others (Cronan et al, 2018). Business school administrators and faculties have developed a heightened interest in cheating by students and in business in general (McCabe & Trevino, 2002; Lawson, 2004; Ballantine, Guo, & Larres, 2016). However, efforts to come to grips with classroom cheating have been inconclusive in part because of a lack of contextual studies and studies examining actual cheating behavior.

In an attempt to better understand the phenomenon of cheating, we examine a situation that captures students' guided reflections on an occurrence of classroom collusion in two sections of an accounting course. Here, we define collusion as secretly sharing information without permission in order to obtain an advantage in the classroom. Ariely (2012) calls this collaborative cheating- where cheating takes place in context of a team. We use Latent Dirichlet Allocation to examine interviews of business students directly involved in a classroom situation where actual student collusion occurred on a relatively large scale. By studying this naturally occurring situation (i.e., where actual collusion has taken place) our results enrich a research stream based predominantly on self-reports of cheating behavior.

LITERATURE REVIEW

The context and morality of cheating and collusion

Most previous research studies have dealt with either the context for cheating or the morality of cheating. For example, McCabe and Trevino (1996: pg. 30) state, "the climate or culture of academic integrity found on campus may be the most important determinant of the level of student cheating on that campus." These authors found that both peer pressure and technology increased the variety of classroom cheating (McCabe & Trevino, 1996). Subsequent papers by McCabe, Trevino, and their colleagues support this conclusion. McCabe and Trevino (1997) also found that peer influence is among the strongest demographic and contextual factors affecting the level of cheating among students. Moreover, at the graduate level, McCabe, Butterfield, and Trevino, (2006) found that perceptions of unethical peer behavior were a justification students often gave for cheating.

Other researchers have also examined the context of student cheating. For example, Kaufmann, West, Ravenscroft, and Shrader (2005) found that students demonstrated immature ethical reasoning and rationalization when they believed peer behavior and the classroom environment encouraged cheating. In responding to open-ended questions about a cheating incident, students revealed both concern with following their perceived norms governing classroom expectations and an ability subsequently to rationalize their behavior. "When everybody cheats, it's okay to join the bandwagon." Another student said, "Coursework is based on the idea of working in teams efficiently and effectively. WE [sic] are so used to this that what others consider 'cheating' to us is 'teamwork'." At the same time, students were not amoral and tried to rationalize their behavior by distinguishing what they had done from what they considered to be more serious, actual cheating behavior. Both the line of reasoning articulated in the studies above and the quotes from Kaufmann et al (2005) suggest that when students perceive an ethical climate in the classroom that is dissonant from their pre-existing moral beliefs or other training, they may behave in ways that are not consonant with these pre-existing beliefs about what is ethically right and wrong.

Another line of research deals with individual understanding of morality and cheating. In the accounting ethics education literature linking moral judgments to actual behavior, Ponemon (1993) describes an extensive effort to teach accounting ethics over a year-long period. Ponemon found that the ethics intervention, as extensive as it was, had little effect on moral reasoning. More interestingly, however, he found free-riding (i.e. not paying the complete cost for study materials) was highest when the moral reasoning scores on the Defining Issues Test were the lowest. Woodbine and Aamirthalingam (2013) also found that Master's of accounting students exhibit very opportunistic behavior and relatively lax attitudes toward cheating when given the opportunity to cheat and advance test scores.

The influence of peer pressure and the moral intensity (Jones, 1991) of the situation may entice students into behaving in ways that others see as unethical (Peterson, 2002; Sierra & Hyman, 2008). As an example, Premeaux (2005) found differences between students at expensive schools with high entrance requirements (tier 1) compared to tier 2 schools. Students at tier 1 schools experience relatively more cheating on written assignments and attached more significant social stigmas to cheating, while tier 2 students, attending local institutions, reported more cheating on exams were more accepting of the notion that even moral people cheat. Similarly, an organizational culture that tolerates widespread sharing of work may precipitate academically dishonest actions. Citing an example where cheating was observed firsthand, Flynn (2003) states that 'displaying concern for one's classmates and seeking to encourage them during an examination... was only natural for someone reared in a culture that emphasized the well-being of the group (pg. 438).' Consequently, organizational cultures may foster divergent attitudes toward cheating, and these different attitudes may create difficulties for course instructors. Such results highlight the importance of student perceptions of classroom context or climate. Therefore, both context and morality affect cheating. Morality

effects are nuanced and related to instrumentality. Context effects appear to be somewhat more direct, that is, students respond to the situation at hand.

The point is that students do not always understand the line of demarcation between cheating and formally assigned group work (Ariely, 2012; Shrader et al, 2012). The widespread use of teams and group assignments may be adding to the level of general confusion about what kinds of sharing are permitted and what constitutes cheating (Kaufmann et al, 2005). For example, a student quoted in the Washington Post article mentioned above (Bever, 2017) said, "I think it [cheating] is not right, but collaboration is required for a lot of classes... so as long as you don't cross this boundary, I think it is good to use it." This statement is fascinating in that it recognizes both context and morality. Some might say it convolutes context and morality. In either case, it indicates that student perceptions of morality and context are important in understanding the phenomena of cheating and collusion (Briggs, Workman, & York, 2013).

Most previous research has examined cheating from an institutional level or has employed survey research to tap into student perceptions. Few have looked at actual situations where students were deeply involved with a particular cheating episode. The purpose of this paper is to present a unique form of analysis of a classroom event where collusion actually took place. In two sections of an advanced accounting course, students were given the assignment of preparing for two sides of a classroom debate. However, some students colluded in an attempt to reduce the amount of work and uncertainty associated with the assignment. We interview the actual students involved with the goal of trying to understand how they viewed the classroom where they either took part in or watched collusion take place. Our purpose here is to seek understanding of this decision to collude by examining what students had to say about the actual situation. Our goal, through the application of latent Dirichlet analysis (LDA), is to identify the underlying structure of words and topics students used to describe their experience. LDA offers the means to explore and provide unbiased insights into the way students actually talk about collusion.

THEORETICAL DEVELOPMENT/MODEL

Study Method

Academic collusion and cheating present researchers with a rather delicate set of research ethics challenges. It would not be ethical for researchers to entrap students into collusion. Moreover, researchers should be extremely cautious about even creating situations where students might be encouraged to collude. Yet few studies examine the background motivations and thought processes of students that collude. The infrequency of actual studies dealing with context belies the fact that collusion seems to occur regularly.

Researchers have relied on a variety of data ranging from student self-reports of cheating behavior (McCabe & Trevino, 1993) to hard evidence such as discarded cheat sheets (Pullen, Ortloff, Casey & Payne, 2000) in order to better understand student cheating. The focal point of these studies varies from the thought process of students who cheated in a particular setting (e.g., Briggss et al, 2013; Kaufmann et al, 2005; West, Ravenscroft, & Shrader 2004; Shrader, Ravenscroft, Kaufmann, & West, 2012), to comparisons between spontaneous and intended cheating (Genereux & McLeod, 1995), to the role of new cheating technologies (D'Souza & Siegfelt, 2017; McCabe & Trevino, 1996).

Notwithstanding the focus or data source, previous studies have come to the general conclusion that cheating behavior is common (Lawson, 2004). More disturbing for those of us who teach in business schools, however, are the comparative studies that have found business students to be among the most prominent among the cheaters (McCabe & Trevino, 1993). For example, in a two-university study involving four hundred students across disciplines, Roig and Ballew (1994) found that students majoring in accounting and finance held the most lenient

attitudes toward cheating among all students in their sample. Compounding this is the finding by Bloodgood, Turnley and Mudrack (2010) that merely taking a business ethics course does not seem to have a much influence on students' views regarding cheating.

Self-reported surveys of attitudes toward cheating, of intentions to cheat, of past cheating, and of reasons for cheating represent the dominant form of data on cheating (Spiller & Crown, 1995; Cizek, 1999, Cronan et al, 2018). While surveys are convenient and useful, and establish the seriousness of the collusion question, they do raise some doubts. There may be a wide gap between how students respond to a survey and how they may actually reflect and respond to questions about their own behavior (Scheers & Dayton, 1987; Karlins et al, 1988; Miceli et al, 1991; Nowell & Laufer, 1997). Consequently, actual behavior provides more validity than self-reports, but presents researchers with ethical challenges because of faculty members' obligation not to encourage collusion.

In this paper, we employ an objective and powerful form of analysis to explore a classroom event where collusion actually took place. We interview the actual students involved in order to tap into their firsthand observations of the classroom dynamics. The sample group is rather small but it represents a major portion of the universe of students involved in the situation. Our intent and method are exploratory. We are not trying to predict collusion. We already know it happened and are now using a powerful form of text analysis to delve into the verbal reasoning of the students about the context. Both colluders and non-colluders were interviewed.

Latent Dirichlet Allocation, Topic Modeling and Words

Current advances in machine learning have led to methods that can be used to uncover hidden or latent relationships present in large data sets. The main difference between these methods and traditional statistical methods is that they make no *a priori* assumptions about relationships present in the data. Latent Dirichlet Allocation (LDA), the topic modeling method used in this paper, is an unstructured machine-learning algorithm that uses probabilistic topic modeling to estimate the likelihood that words are grouped in similar topic areas (Zupic & Cater, 2015). LDA is an analytic technique where observations in a large set of data are explained by similarities in otherwise unclassified groups (Blei, Ng, & Jordan (2003). LDA extracts patterns from a large set of data without a priori assumptions so it is not designed to answer specific pre-formulated hypotheses (Schwab & Zhang, 2018). Instead, it is an exploratory technique, one that provides fundamental word-level observations on basic language structure found in a large volume of text. In this way, LDA opens up large data sets to unrestrained discovery. It is a method of discovery that Moro, Cortez and Rita (2015) claim is useful for conducting progressive and relevant research in any disciplinary field. Implementations of LDA and its variants are available from data analytics firms, academic software repositories, or from open source software such as Github.

As an unstructured machine-learning algorithm, LDA makes no assumptions in terms of how information sorts into silos. A researcher inputs textual data to the algorithm, which produces words and topics from the data. The richness of LDA is that it recognizes that there can be many probabilistic topics and that words are independently distributed among topics (Sugimoto et al, 2011). In our exploratory study, student interviews are the corpus (data set) from which we specify words based on probability distributions called topics.

The basic unit of modeling is the topic. Topics are combinations of words. A topic is a distribution of words over the entire set of words in a corpus (Sievert & Shirley, 2014). Most LDA applications focus on textual data because LDA is able to analyze the underlying structure of large amounts of text (Sugimoto et al, 2011). It is the most widely used topic modeling method because it allows researchers to extract a parsimonious set of an optimum number of latent dimensions or topics in collections of data (Sugimoto, Li, Russell, Finlay, & Ding, 2011; Tirunillai

& Tellis, 2014). Dyer, Lang, and Stice-Lawrence (2017) use LDA to ascertain exactly “what is being said” in corporate reports. We use LDA in this study to look for the most referred words in a set of student interviews. In LDA, a ‘word’ is the basic unit of discrete data (Blei et al, 2003). A ‘document’ or interview is a sequence of words, and a ‘corpus’ is a collection of interviews. Specifically, we employed LDAvis (Sievert & Shirley, 2014), an LDA-based data visualization method, to analyze our data. LDAvis allows us to examine how different words contribute to the meaning of each topic, how different topics relate to each other, and the prevalence of each topic. Our corpus is the entire set of comments transcribed from the interviews with the students involved in the colluding and non-colluding class sections. The corpus includes the text from twenty-six student interviews.

In effect, LDA helps discover underlying themes in a set of data by generating key words. Words are allocated to topics in the analysis. As domain experts, the researchers in effect, interpret the topic results. In our study, the domain experts are both the authors and the LDA administrators. Therefore, a topic is a cluster of words specified by the domain experts. Unimportant words (e.g. a, and, are, is, the etc.) are often ignored or reduced in the analysis. Rather, the important key words, words that are idiosyncratic and exclusive to a topic, are identified and emphasized.

LDA does not create new words or concepts as with factor analysis or content analysis; rather topics in a set of data are represented by existing key words (Moro et al, 2015). In a well-formed topic model, certain topics will generate words from one conceptual area more than from another - for example, “repercussion-related” and “team-related” topics, where the former is more likely to generate words like “caught”, “colluded”, “decision”, “fail” and “trust”, and the latter words like “sharing” and “teammates.” Therefore, if an interview contains primarily important words, the model will classify them by topic. For each interview, LDAvis determines the probability of the set of words in each interview belonging to a topic, and then it matches the words with topics. The LDAvis relevance metric sets the weight given to the probability of a word belonging to a topic (Sievert & Shirley, 2014). An interview may represent different topics. Topics are based on probabilities assigned by the LDA software.

In LDA, the administrator determines the number of topics. The appropriate number of topics is determined in many ways (Chen & Wang, 2018). Too many topics dilutes the meaning of each topic and too few is not discrete and does not separate ideas and words from each other. The appropriate number of topics for our dataset was determined by comparing the intra-topic similarity with inter-topic dissimilarity. We ran five, ten, and fifteen topic solutions. and chose a result that maximized the difference between topics (Chen & Wang, 2018). We determined the optimum number of topics to be five and the most illustrative relevance metric to be .6 (as recommended by Sievert & Shirley, 2014). Our LDA produced for each of the five topics a list of the most relevant words. Relevance of the word w to topic k given λ is defined as:

$$r(w,k/\lambda)=\lambda \log(\varphi_{kw}) + 1-\lambda \log(\varphi_{kw} / p_w) \quad (1)$$

where φ_{kw} is the probability that word w belongs to topic k and p_w is the marginal probability of word w being in the corpus. Lambda can be set to values between 0 to 1. A lambda of 0 would equate relevance to exclusivity – where the top words would be the ones whose probability of being in the whole corpus and of being in the topic are the same or very close. A lambda of 1 equates relevance to the probability of being in that topic, which is the frequency of the words divided by all the total number of words in the topic. As recommended by Sievert and Shirley (2014) who developed the relevance measure, we use a lambda of 0.6 as noted above.

Based on each topic’s most relevant words, we decided on topic names. We labelled topics by determining the mix between the probability and relevance of words belonging to topics (Sievert & Shirley, 2014). Our topic results and labels are based on output directly from the analysis, and represent the underlying themes students use to describe their personal experiences with collusion. Therefore, LDA emphasizes topics and words. Words structure the

topics. We focus on words in particular because words represent the most subtle and yet critically important way to delve into the understanding and assumptions that made their way into student thought and action. Words provide a glimpse into the mind and will of the students as they describe the collusion incident.

Prior to performing this analysis, we had information that separated colluders from non-colluders. In the interviews, we expected that there would be clear differences in the question responses between colluders and non-colluders. However, our objective in using LDA was to engage in a more granular examination of the language or topic differences, if any, between the two groups. Therefore, we ran the analysis of the entire set of interviews as one corpus. We considered the colluders and non-colluders responses together in order to objectively examine the language. LDA is bias free in terms of caring what a colluder or non-colluder says. It simply accumulates words into clusters or topics and provides for a visualization of the differences. As part of this analysis, we attempt to examine the topic differences among individuals. We trace individuals to topics in an attempt to understand how colluders and non-colluders might be different.

Zupic and Cater (2015) state that LDA holds tremendous potential for “expanding the scope of mapping the management and organization domain (pg. 457).” In order to make wide application of LDA available in organization research, Zupic and Cater indicate that ethics scholars can either wait for new software to be developed or collaborate with information scientists. In our case, we interacted with technical experts from a data analytics firm- Kingland Systems. Kingland, headquartered in Clear Lake, Iowa, provides information technology and financial services to manage compliance and risk for firms. Company clients include some of the world’s largest banks, financial services firms, and insurance companies. We worked directly with Kingland analysts, providing them the raw data, or corpus, and then collaborated with them to interpret LDA results.

Data for this study are derived from the complete set of interviews for all student participating in the interviews. We base our LDA on the analysis of the entire interview text, including the initial greeting, general conventions of opening a dialog, and the entire set of responses to the questions posed in Appendix A. This method is analogous to the corporate implementations of LDA used by our partner, Kingland Systems. For example, Kingland provides their clients full-text analysis of formal corporate reports and compliance documents. The authors of this paper interviewed the students and generated the transcripts. Kingland analyzed the transcripts and produced the LDAvis results. The Kingland experts directly helped the authors ascertain the number of LDA topics to best analyze the corpus and they offered advice on the relevance metric to best identify topic words. We use all the text of the interviews in our analysis, thus providing a complete unbiased corpus for analysis.

Sample- The Collusion Incident

A strength of this study is the sample. It is in effect, a naturally occurring field experiment. We did not manipulate a cheating or collusion intervention. It would be unethical to do so. The collusion occurred naturally. Therefore, the strength or external validity of the experimental effect is greater than would normally be expected with a lab experiment or from a survey. We were able to obtain access to the students involved through the course instructor.

The incident we discuss and analyze occurred in a graduate-level course on accounting professionalism and leadership, taught at a large, public university in the Midwest. Most of the graduate students in the section were accounting/tax majors. A major, multi-part assignment of the course was a debate tournament. The faculty member involved professionals from the nearby metropolitan area to observe the debates and to offer their evaluation of the students' performance. While their presence added a connection to the "real world", the professionals added to the anxiety students felt about doing well on this assignment. The degree program the

students were enrolled in is known to be intense and this course was considered very time-consuming.

The debate assignment was designed to develop multiple skills. Research and critical thinking and analysis were required to formulate a strong response to the ambiguous and technical accounting questions. Additionally, the professor wanted students to practice speaking publicly after having carefully prepared their main arguments, but also facing the challenge of speaking extemporaneously if the opposing side created new arguments or offered new evidence. Of course students could mitigate the likelihood of encountering new evidence or novel arguments by preparing very intensively themselves. To ensure students understood the issues thoroughly, the debate process required that teams flip a coin before each debate and the winning team could select the side of the issue they preferred. That uncertainty about which side they would be arguing caused students to spend even more time preparing.

Students did know which team they would be arguing against before the coin toss. However, they were told explicitly not to collude with their opponent to predetermine debate sides. The University has an honor code that mandates both that students refrain from cheating and that they disclose incidents of cheating which they observe. The instructor represented to the external judges that groups would not know which side they would debate until the coin toss.

Before the last round of the debate, the instructor became aware that some teams had colluded with the opposing team to predetermine which side each would argue. The professor overheard two teams deciding, despite the professor's proscription, to work together and collude to pre-determine which side they would argue. After the debates concluded, a student (who had explicitly declined the possibility of collusion) reported to the professor that the collusion was more widespread than he had first thought.

The professor reacted strongly to the collusion for several reasons. Because of the collusion, he felt he had misrepresented the debate to the professionals who had volunteered their time and expertise. He believed (as did the students involved) that colluding provided a competitive advantage over teams that did not collude. More importantly, collusion was blatantly unfair and constituted a form of academic dishonesty and violation of the school's honor code. On a personal level, he felt disrespected; furthermore his goals for the class had been stymied and the deep rapport he thought he had with the students was dissipated. However, he wanted to understand why some students had chosen this path, and he wanted the students to learn from the experience. Therefore, after confronting the class, he approached us about interviewing them about the incident.

Table 1 includes student comments made directly to the instructor regarding the collusion incident. In these comments can be seen some of the word topic themes that appear in our formal analysis. Student acknowledged that they felt team and peer pressure to collude. There also appear to be comments related to the idea that the students understood that they were being dishonest and that there would be repercussions for their actions. A majority seems to be remorseful for their behavior. However, the students also offer rationalizations that collusion helped lighten the perceived high-pressure workload. The corpus for analysis includes the text from twenty-six student interviews. The twenty-six students volunteered to be interviewed - no one was compelled. Eighteen interviews were with non-colluders and eight were with colluders. Of the students interviewed, 12 were female. Of the eight colluders, three were female. Therefore, there was gender balance in the sample.

Interviews

The interviews produced 524 pages of text, and while not large by typical LDA standards, it does represent the entire set of comments about this incident. There were 167 pages of text collected from the colluders, 339 for non-colluders and 18 for the outside observer.

This means that each colluder provided approximately 21 pages of text and each non-colluder provided approximately 20 pages.

Student subjects were interviewed via telephone by the study authors. Each interview was conducted with permission of the respondent and the home university's Institutional Review Board. The interviews were immediately transcribed by an independent service and the text of the interviews constitutes the corpus for this research. We asked each respondent a similar set of questions about the collusion incident. The questions are presented in appendix A.

We examined the complete text of all interviews that included 11,125 paragraphs and 218,993 words. There were 65,439 total words for the colluders and 153,554 words for the non-colluders. This results in approximately 8180 words per colluder and 8531 per non-colluder. Given that the responses were freely offered and not constrained, non-colluders were slightly more verbose. Non-colluders talked more about the experience.

Results

The results indicate the words most associated with the five interview topics. Topics in this case represent the substance of the interviews. The two visual features in figure 1 provide overall perspective on these five topics. Circles identify the topics plotted in terms of multidimensional scaling using LDAvis. The size or area of each circle indicates the proportional prevalence of topics in the corpus (Sievert & Shirley, 2014). Three topics (2, 4, and 5) are overlapped and comprise 76.1% of the entire corpus. Overlap simply means the topics are related. Topic 4 is the most prevalent in the corpus (40.7%). Topics 3 (19.1%) and 5 (23.9%) are also prevalent. Topic 3 is interesting. It is not overlapped but is near the other overlapped topics. It is rather important, however, in that it represents 19.1% of the corpus.

The other visual feature in figure 1 lists the top most probable or most salient terms or words from each topic cluster. The topic clusters all contain non-specific and common words, such as 'yeah', 'like', and 'okay.' This is an expected outcome of LDA. Each topic, however, also contains words that are more idiosyncratic or unique to the topic made manifest through the relevance metric. These words cluster to a given topic exclusively compared with the more common terms. We used the word clusters to describe each topic in Table 2 following convention practiced elsewhere in the LDA literature (e.g. Sugimoto et al, 2011; Lash & Zhao, 2017). A complete set of LDAvis diagrams is presented in appendix B.

Table 2 includes the topic description, percent of the corpus represented by each topic, and the most relevant and unique words. The authors determined the topic descriptions based upon relevant words and our understanding of the situation. These relevant words provide the bulk of our analysis. As noted above, relevant words are prevalent in the topic within and throughout the corpus. Unique words are those prevalent in one topic at the exclusion of others. Uniqueness was determined simply as the percent of a word in a topic, that is, the prevalence of a word in a topic divided by the prevalence of the word in the entire corpus. For example, in topic 1, the word 'interview' has a uniqueness score of 99. This means that even though the word can appear elsewhere in the corpus, there is a 99% probability it will be associated with topic 1 in the LDA.

Table 1	Sample post-collusion student comments made directly to instructor regarding the collusion
Student #1	"I wanted to inform you about my side of the collusion mishap. For the third debate only, about an hour before class, I was approached and asked if my team would take the Con side because one of my opponents was not prepared. I said yes because I felt bad and I didn't want my friends to be embarrassed. I realize that it was a bad decision and I wasn't happy about it at the time. I made a bad decision due to pressure and stress. Although it is not an excuse, my team was fully prepared on both sides of the debate. I am taking this situation as a learning experience about how easy it is to fall into unethical situations."
Student #2	"I would like to tell you that my group only talked with the other groups about what side they preferred and then we flipped a coin to find out who would debate on what side. After that we parted our separate ways and didn't speak any more about the debate, the contentions or the evidence. Nonetheless, we still researched both sides of the issue, and prepared good contentions as well as provided strong evidence to support them. If you would like I can send you our contentions as well as our evidence to support them for both sides. It was never our or my intent to be disrespectful to you, the guest judges, or other members of the class. I feel that I learned a lot from this experience and will be more thoughtful about my actions when working on projects."
Student #3	"First off, I know without a doubt that I was one of the people you overheard conversing with my opponent the day of the debate. As I look back, you probably watched me do it. On the final debate, there was an arrangement with our opponent that we would take the con side if we won the toss, and our opponent would take the pro side. What I am most ashamed of is that I knew it was wrong, but I did nothing to stop it. It is also wrong to rationalize it. I'm not going to blame my partner or say, "Well the other teams are doing it". Yes, it was a difficult week and I had a lot on my plate, but there is no excuse for what I did."
Student #4	"The bottom line is that I panicked about 1 hour before the debates. I was personally extremely well prepared for both sides (seriously, I'll show you an epic amount of research organized in two folders pro and con), but my partner was not. I knew most teams strongly favored the con side, so I pushed him to focus on the pro. Additionally, in previous rounds I had been approached by other teams to pick a side days before, and I declined."
Student #5	"I was a participant in the collusion! To be honest, I am pretty embarrassed about the situation. I took the approach to work smarter, not longer or harder during that week. Not very proud of it, but I cannot go back and change it."
Student #6	"I am sorry to say that my team had our side predetermined in the final debate. Although we did not end up debating, I definitely regretted my decision after realizing the magnitude of what we did. I apologize for destroying the whole point of the debate, and I have no excuse that would validate it."
Student #7	"I am responding to your message because member of my team was involved in the collusion. I had mentioned some of this in my final memo so part may seem repetitive if you have already read it. He made an arrangement with one member of the team we were supposed to face. He then sent a text stating what he had done. It was done without my knowledge or approval. I believe he did it to spite you or the debate. I think it was like you said in the email, he was overwhelmed by the work load and was doing what he could to make the work load manageable."
Student #8	"With heavy heart, I am admitting that group one participated in the collusion. We flipped the coin the day before debate. As ashamed as I and my teammates are, I don't feel it is appropriate at this point to try and justify our actions besides agreeing with what you put in your email. I am truly sorry for the disrespect, and for letting you down."

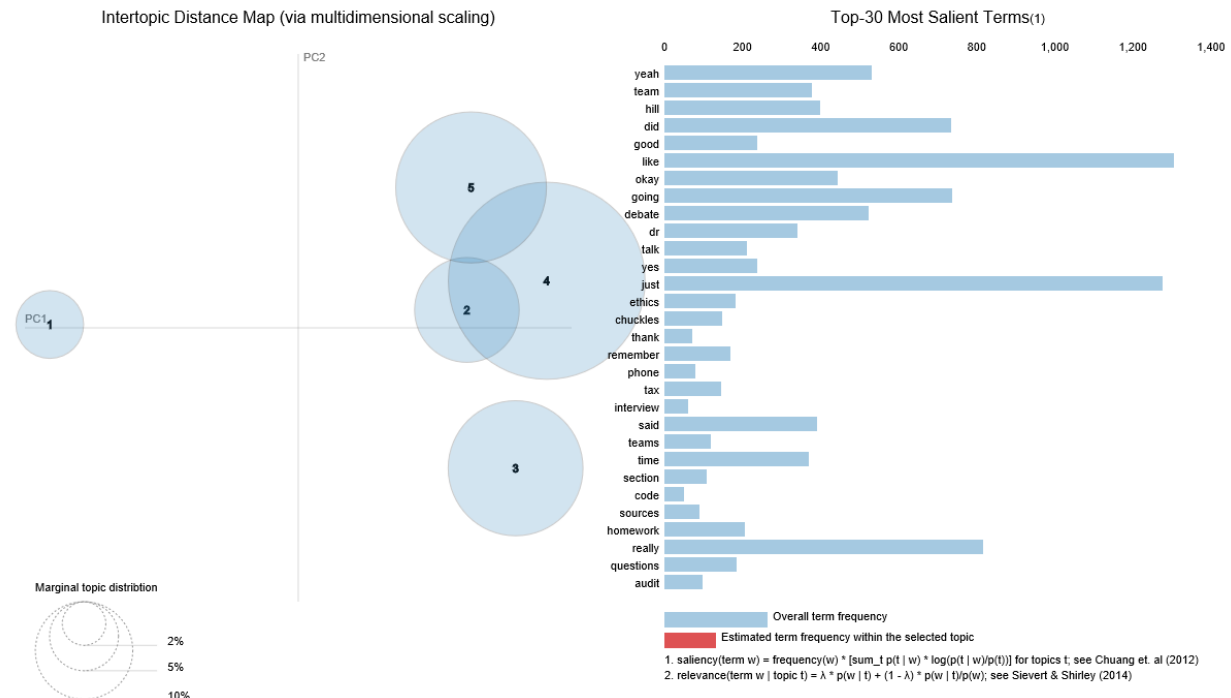


Figure 1- Topic map of student interviews

Evidence from the other high-loading relevant words reveal substantive topics dealing with the nature of the course assignment (topic 2), class teams (topic 3), repercussions of being caught (topic 4), and differences between the auditing and tax tracks in the accounting major (topic 5). Topics 2, 3, and 5 constitute over half of the corpus (55.4%). These topics include words related to taking sides in the debate, teams, and the differences between the auditing and tax majors. Therefore, in effect, students talk most about the immediate classroom situation when describing collusion. Topics 2, 4, and 5 cluster together and deal with class competition, taking sides, repercussions, and differences between majors. Topics 2, 3, and 5 deal with

Table 2 – Extended latent Dirichlet allocation results for student interviews

Topics and Words

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	
Percent of Corpus	4.8%	11.5%	19.1%	40.7%	23.9%	Total = 100%
Topic Label	General context- honor code, starting the interview, naming	Nature of class assignment itself	Teams - a resource and support mechanism	Repercussions of being caught	Tax and Audit class section ethical differences	

	important actors					
Important Words	<u>Relevant words (LDavis)</u> thank interview code phone yeah hi hello	<u>Relevant words (LDavis)</u> like going just really said article prepared	<u>Relevant words (LDavis)</u> team did debate remember know teams groups	<u>Relevant words (LDavis)</u> think class people know students caught colluded	<u>Relevant words (LDavis)</u> think ethics just people know tax audit	
	<u>Unique words (% of word unique to topic)</u> interview (99) code (99) honor (99) hello (96) codes (94) interviewer (93) honor (93)	<u>Unique words (% of word unique to topic)</u> confident (99) article (98) argument (97) argue (97) articles (96) pros (96) cons (90)	<u>Unique words (% of word unique to topic)</u> sources (99) shared (98) competing (98) agreement (95) teams (94) teammate (80) team (67)	<u>Unique words (% of word unique to topic)</u> dishonest (99) fail (98) recommend (98) hiring (98) failed (97) determining (97) probability (96) failing (95) stupid (93) caught (92)	<u>Unique words (% of word unique to topic)</u> ethics (99) tax (99) emphasis (99) engaging (98) preconceived (98) culture (96) curriculum (96)	

things more internal to the particular collusion incident while topic 4 deals with the external view of the reputation of the school. This appears to be evidence that internal course context has a lot to do with collusion. Table 3 provides samples of student interview comments that reflect the topic categories.

Topic 1 – General context, honor code and greeting

As evident in table 2, the relevant (common) words for topic 1 are thank, interview, code, phone, yeah, hi and hello. Words unique to the topic include interview, code, codes, interviewer, honor, thank and hello. Therefore, topic 1 captures language regarding the basic greeting and simple pleasantries associated with opening the dialog between interviewer and interviewee. It is also quite separate from the other topics on the inter-topic distance map and is much less prevalent than the others (4.8%). Along with this, it also appears to deal with the school code of ethics or honor code. This school did indeed, have an honor code and the interviewees seem to mention it as part of the general context of the situation. McCabe and Trevino (2002) argue that academic integrity is enhanced through a visible honor code. Here we find the code to be visible and part of a general context. The word 'code' does not figure heavily in the more prevalent topics.

Table 3 - Sample student interview comments

Topic		
Repercussion	Non-colluder	"Before the debates started, the professor was very adamant to our class that it was not an option and if we were caught-- I don't know if he said the consequences would be bad, but it was one of those things where it would not be tolerated and personally to him, he would be very disappointed in us."
	Colluder	"I'm aware that we're cheating and being dishonest, so trusting people to keep their mouth shut is a little difficult. If they're dishonest on one thing, who's to say they're not going to be dishonest on another thing?"
Team	non-colluder	"I just think our class had a lot of students that were very talented, and I think we had people that wanted to succeed and were competitive. I think when you combine how you have a lot of talented and competitive people then it brings everyone up to that level of wanting to compete. There was definitely some rivalries between students as well; there was one team that me and some people that I knew in the class were really wanting-- hoping they would lose."
	Colluder	"Teams would give other teams the references and show them where to find the things. I was thinking that we have to trust each other and we have to be quiet about it."
Nature of assignment	Non-colluder	"Yeah I think there is a little difference. It's based on the seriousness of the debate compared to just one homework assignment."
	Colluder	"We were in front of our classmates, we didn't want to look stupid. When the third one came around, we were just so exhausted from everything, we thought it was actually our right to cheat or make it a little easier for us to fix the third debate. We felt we had so much on our table that it was justified to just pick one side, because we have to do the debate anyway."
Audit and Tax section differences	Non-colluder	"As weird as this is going to sound, the class that I was in consisted most of auditors, and we have skeptical minds to begin with, so we are a little more distrusting with each other. One of my concerns with even colluding would have been, "Hey, are they going to come back and throw me under the bus? If I say I'm going to be pro and they're going to be con, are they going to pick pro and I'm not going to have anything to argue with?" For our class we have a more competitive mindset than the other class. We're very close and talked about a lot of [?] and just-- they worked together on a lot of things. We were a way more competitive group than they were."
	Colluder	"Tax is the outcast of the accounting program. Usually people go into audit where it's more regulatory stuff. The tax people are sort of-- I don't know, it's different. Everything in the audit program is very structured and you have to say the right words, or it doesn't make sense. They do more writing things and analysis. The tax people are more transactional. We have to be more outgoing with clients. We actually have to talk to our clients, and the work that we do is start to finish. We start a tax return, we finish it. An audit, you're auditing cash. You're only one segment of the big picture. It seems like tax people are more lawyers, we deal more with like cases and interpretation. Whereas an audit, it's more concrete and it's not just the same. Honestly, personality decides who's going to be an auditor and who's going to be a tax person."

Topic 2 - Nature of course assignment

Topic 2 is the second least prevalent topic in the corpus. The most relevant words are: like, going, just, really, said, article and prepared. When combined with unique topic words such

as: confident, article, argument, argue, articles, pros, and cons, it is clear that this topic deals with the course assignment of developing an argument for the debate.

Topic 3 – Teams

Topic 3, which deals with the team aspect of the course, appears to be somewhat independent of, but related to, the other main collusion topics in terms of the distance map. This may mean that the team aspect itself has a lot to do with the decision to collude, or at least teams create relevant context for the decision. The finding that 'teams' is a topic unto its own illustrates the importance of teams and teamwork in understanding cheating and collusion. Important relevant and unique words include- shared, agreement, sources, teams, competing, teammate, team, remember, and groups. These words clearly highlight the salience of the team concept in this situation.

Topic 4 – Repercussions of Collusion

The most prevalent single topic in the corpus (39.3%) is topic 4. The words fail, failed, failing and caught all led us to describe this topic as having to do with the repercussions of collusion. Both the immediate repercussions (fail and caught) and future consequences (future, recommend, and hiring) appear to be important considerations to the students. What is interesting is that even though repercussions of being caught are prevalent, many students were still willing to take the risk. This result follows that of West et al. (2005) who found students to become very instrumentally focused when faced with opportunities to cheat. In other words, when the payoff is large enough, even severe consequences many not deter the colluder.

Topic 5- Difference between audit and tax classes

The words 'tax' and 'audit' are both prevalent and unique in topic five. These words describe the nature of accounting majors in different sections of the course- one populated by tax majors and the other audit majors. Other important words are ethics, curriculum, engaging, emphasis, and culture. The collusion incident happened in a section populated by tax majors. It is clear that these students perceive the audit and tax people to be different. In this instance, the audit class did not have a problem with collusion. One explanation for why auditing students may be less susceptible to collusion is found in the work of Ballentine, Guo, and Larres (2018) who demonstrated that less cheating was found in instances where students were involved in deep and strategic approaches to the material. Tax laws and rules may concomitantly reflect what Ballentine et al, term surface learning approaches- requiring memorization without deep understanding. Whatever the reason, it is clear that in this situation involving collusion that the audit and tax sections were seen as different, and though surface versus strategic learning are not part of our research, the conditions appear to be consistent with our findings.

Colluders, non-colluders, and topics

We also extend the LDA results to individual students. Table 4 presents information linking individual students to LDA topics. The general pattern reflects the topic relevance in the corpus as expected. Concern regarding repercussions, topic 4, was the most relevant topic for a vast majority of those interviewed. Topic 4 was not the most relevant topic for only two of the student respondents. There were only subtle differences between colluders and non-colluders in terms of relevant topics overall. However, according to the Kingland LDA experts, 'even subtle differences are important and interesting.' The difference between the audit and tax majors and the nature of the class assignment were slightly more relevant to colluders than non-colluders, while teams were marginally more relevant to non-colluders. Also, any names appearing in the list of terms are disguised and, therefore, do not violate anonymity.

**Table 4- Extended LDA results
Most Relevant Topics by Individual Colluder/Non-Colluder**

Colluder/Non-colluder	Most Relevant Topic	Second Most Relevant Topic	Third Most Relevant Topic	Fourth Most Relevant Topic	Fifth Most Relevant Topic
Colluder 1	4	5	3	1	2
Colluder 2	4	5*	3*	2	1
Colluder 3	3	4	5	2	1
Colluder 4	4	5	3	2	1
Colluder 5	4	5	3	2	1
Colluder 6	4	5	3	2	1
Colluder 7	4	3	5	2	1
Colluder 8	4	5	2	3	1
Non-Colluder 1	4	3	5	2	1
Non-Colluder 2	4	3	5	1	2
Non-Colluder 3	4	3	5	2	1
Non-Colluder 4	4	3	5	1	2
Non-Colluder 5	3	5	4	1	2
Non-colluder 6	4	5	3	1	2
Non-Colluder 7	4	5	3	2	1
Non-Colluder 8	4	3	5	1	2
Non-Colluder 9	4	5	3	1	2
Non-Colluder 10	4	5	3	2	1
Non-Colluder 11	4	5	3	1	2
Non-Colluder 12	4	3*	5*	1	2
Non-Colluder 13	4	3	5	2	1
Non-Colluder 14	4	5	3	2	1
Non-Colluder 15	4	3	5	2	1
Non-Colluder 16	4	5	3	2	1
Non-Colluder 17	4	5	3	1	2
Non-Colluder 18	4	5	3	2	1

**topics tied for this individual*

Conclusion and Discussion

To our knowledge, this is the first time LDA has been applied to a dataset to examine cheating or collusion. Evidence from our LDA suggests that a significant portion of the

language used to describe collusion is manifested by four major topics. Students see repercussions of being caught, teams and class teamwork, course assignment structure and content, and class section differences to be important factors in the context of collusion. Indeed, the most salient topic in this research deals with repercussions. Our LDA results align closely with those of Briggs et al, (2013) indicating that collusion is a rational assessment on the part of students guided by the expectation that other teams will collude and fear of exclusion on an assignment, even when harsh repercussion may be applied. The discussion also centers on differences between auditor and tax functions in accounting. Taking the repercussion and auditing difference topics together coincides with previous research that finds auditor moral reasoning to be influenced by the potential for penalties (Jeffrey, Dilla, & Weatherholt, 2004). Repercussions as an assessment of the instrumental nature of decisions to cheat are common elsewhere in the literature (West et al, 2004).

Another of the topics resulting from our analysis centers on teams and teamwork. Our results suggest that collusion may be associated with team assignments. These results also confirm the survey results of Shrader et al. (2012) that teams and peers may put unintentional pressure on students to collude on class assignments. Teams were part of the class structure in this incident as they are in many business classrooms today. In business schools, we often use teams to teach consensus building and appreciation of diversity. Instructors place students in teams hoping to maximize individual performance and enhance inclusion. However, teams also can create blurred lines. Teams often result in shared work- both when desired and when not permitted. Team assignments may lead to free riders both within the team and including the whole team. Team based learning is a popular teaching technique, but using teams in the classroom requires careful structuring of activity and the team itself in order to be effective (Koppenhaver & Shrader 2003). It appears that whenever teams are formed there is an enhanced opportunity for students to use the team in an instrumental way to lighten the workload. As Ariely notes: "Some...forces might make it easy for group-based processes to turn collaborations into cheating opportunities in which individuals cheat to a higher degree because they realize that their actions can benefit people they like and care about (2012, pg. 222)." Ariely claims that under conditions where teammates perceive they can all benefit from the opportunity, the level of collaborative cheating tends to increase. The implication for future research is that team context must be considered because of its powerful effect on class assignments. The widespread use of classroom teams for presentations and debates may be creating confusion (or opportunity) in terms of what is allowed collaboration and what is illegal collusion. Forming classroom teams without providing team training and time for team development may be doing as much harm as good. Certainly team-based learning is more effective if teams operate as planned. One of the central issues of team-based learning is the degree to which the instructor may have influence on the teams. However, collusion and the subsequent conflict that commonly occurs is likely to have a negative impact on learning, and can minimize the effectiveness of classroom team assignments.

Our results also show that those who chose to collude and those who did not described the incident in similar terms. When asked a battery of detailed questions, there was no glaring difference in terms of how students described the incident. Often students may not see the line of demarcation between cheating and formally assigned group work (Shrader et al, 2012). The widespread use of teams and group assignments may be adding to the level of general confusion about those specific instances when sharing is permitted (Kaufmann et al, 2005). For example, the student quoted in the Washington Post article mentioned above (Bever, 2017) sees no problem with taking advantage of opportunities to collude when not expressly prohibited.

One fundamental aspect of this paper is that we demonstrate the value of large data processing techniques such as LDA and LDAvis. LDA provides an objective and consistent

means of examining large amounts of text. It allows the identification of underlying content and discloses words and topics to make the content interpretable. Our colleagues at Kingland use it to analyze volumes of corporate annual reports and technical documents. It is well suited to this task. However, to our knowledge, this is the first LDA study to examine the language of students who are actually involved in a classroom collusion incident. Moreover, even though not large, our sample represent the universe of data available on this particular collusion incident. Therefore, there is clearly an opportunity for future LDA researchers to systematically gather interviews from larger groups of students over time in order to ascertain even better understanding of cheating and collusion.

Consequently, our study is subject to the basic caveat of using LDA, namely, topics require interpretation by the researchers. The number of topics is subject to interpretation- it is not a necessary given from the analysis. However, we followed the conventions of prior LDA studies in terms of identifying topics, and used LDAvis to convey results. We also worked closely with the Kingland LDA administrators in deciding on the topic solutions and interpreting results. Therefore, we feel confident our interpretations are reasonable and accurate with regard to our data. So, what have we learned from LDA? We learn that students view classroom cheating from the classroom level. Rather than talk in terms of greater ethical principles or codes, they talk about instrumentalities and temporal immediacy. The nature of the assignment, the team dynamic, and the perceived immediate consequences weigh more heavily in their language than do big picture long term things. Students view the decision to collude as a function of course assignments and perceived workload. Their moral reasoning seems to be based on the here and what near-term consequences might be. The immediate context is more relevant to them than the bigger picture. Given this, our research fits with previous research indicating that both context (e.g. codes, McCabe & Trevino, 2002) and morality (West et al. 2005) matter. Our study found only small differences between how colluders and non-colluders describe a shared experience. Therefore, while not expressly tested, our results appear to be consistent with this general finding.

This is also the first time of which we are aware that unique words have been used to identify and discuss topics. By identifying unique words, we were, in turn, able to more clearly identify topics. LDA is free from a priori assumptions about the nature of concepts or factors. The identification of unique words aids in the exploration of the corpus and naming of topics, in our view. Uniqueness is used in this paper to refine, not define, the topics. It offers a more granular statement about the meaning of the topic and helps clarify the results of our exploration.

A weakness of this study is that the corpus is not large. The number of interviews and amount of text is not particularly great in terms of what most LDA studies examine. However, we do have all the possible or available interviews. Our sample is actually a universe and is, therefore, representative. We have all the possible data from this naturally occurring filed experiment. We believe it very reasonable to assume that our students were telling the truth in the interviews and that the interviews captured the essence of the collusion issue. We also view the accounting course in question as typical of other graduate accounting courses and that it created classroom dynamics similar to those of other business courses taught elsewhere.

However, even with these limitations, this research has practical teaching implications. Our results clearly indicate that the course context is salient in terms of collusion. Future research should delve more deeply into group and team issues, that nature and clarity of the course assignment, and the extent to which students will compare themselves with other in terms of the cheating opportunity.

APPENDIX A – Phone Interview Questions

1. How did the idea of colluding in general first come up? What made you think that this was an option?
2. Do you know who the first group to collude was?
 - a. How did you find out about this group?
3. Did knowing that other groups were colluding influence your decision to collude? If so, how?
4. Who first brought up the idea of predetermining sides in your debates?
5. Do you believe that there would be any ramifications to saying no to a classmate who wanted to collude on the debates? If so, please elaborate.
 - a. What about if a classmate wants to look over your homework before it is due?
6. Do you believe that students in previous sections of this course predetermined sides in the debates?
 - a. - If so, why do you believe this?
 - b. If previous sections engaged in the same behavior, do you have any feelings about the fact that you were discovered and they were not? If so, what are those feelings?
7. Has anyone taking the course this semester talked with you about the debate process?
8. Do you believe that NIU's emphasis on ethics had an impact on the debate situation?
 - a. If so, what? If not, why not?
9. Do you believe that there are repercussions at NIU for engaging in unethical behavior such as cheating?
10. What did you think would happen if you were caught predetermining sides in the debate?
11. Is there some difference between the classes that would make one more likely to predetermine debate sides than the other? If so, what?
12. Why do you think that the students in the other section did not collude?
13. Did the fact that the majority of the class was on the tax track and taking all of their classes together have any influence on the decision to collude? If so, can you describe that influence?
14. How do you think that members of the other class would feel or react if they knew that you had colluded on the debates?
 - a. Do you know if anyone in the other section knows that you predetermined sides?
15. How stressed were people during the debate process?
 - a. How did they deal with the stress?

16. What role do you think the debate stress played in people's decision to collude?
 - a. (if significant) What was the difference between the students who were stressed and colluded and those that felt stress and did not collude?
17. How clear was Professor Smith in letting you know that predetermining sides in the debate would be considered cheating?
 - a. Did Dr. Smith say anything before the debate about agreements between competing teams?
 - b. Did Dr. Smith say anything before the debate about working with people outside your team on the debate?
18. How important was trusting the members of the other team on your decision whether to collude?
19. How prepared would you and your team have been if you had been double-crossed and your competition chose to argue the side that you had prepared?
 - a. How confident were you that this was not going to happen? Why?
20. Is there anything about Dr. Smith or this class that would make students more willing to engage in collusion or similar activities when compared with other classes or other professors?
 - a. Debate process itself?
21. How is this class viewed by the students?
 - a. How does this class fit into the Accounting program?
22. What do you think about how Dr. Smith handled this situation after discovering the collusion? How do you feel about it?
 - a. How would you have handled it if you were him? Why?
23. How well known is this response within the program?
24. Do you believe that there will be any repercussions associated with this response?
 - a. e.g., on future cheating?
25. Do you believe any of teams shared evidence or sources with other teams during debate preparation?
26. Is there a difference between discussing arguments or sources with another team in the competition and discussing these things with the team you are competing against? If so, what is the difference?
27. Is there a difference between helping your friends / classmates on a homework assignment that is to be graded and colluding in the debates? If so, what is the difference?
28. Do you believe that your professors are okay with sharing homework assignments that are to be handed in for a grade?

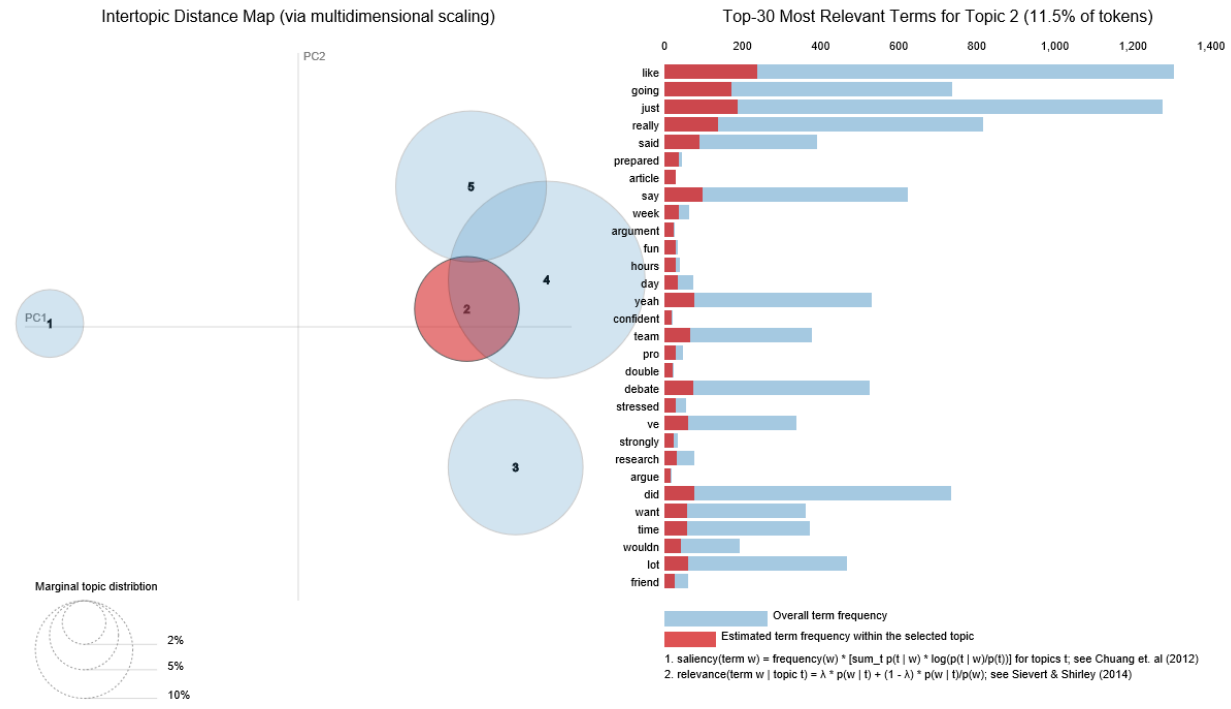
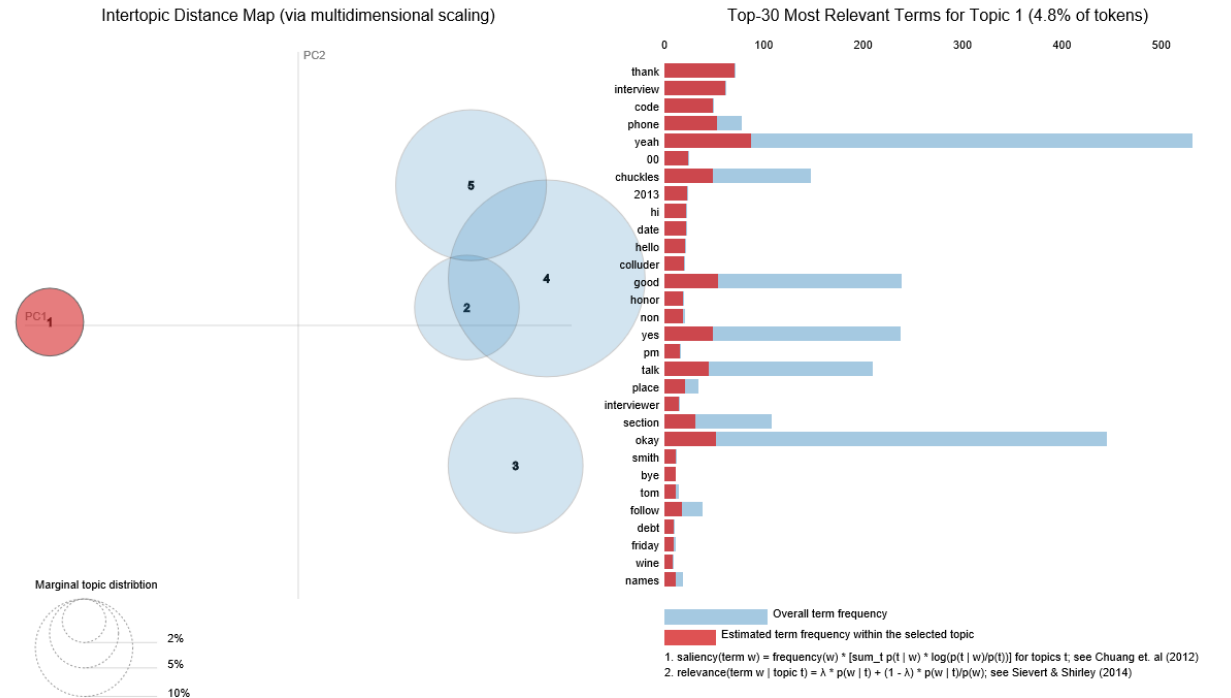
- a. if yes, would you tell them you did it or keep it to yourself? Why?
- b. If no, how does sharing homework differ from predetermining sides in the debate?

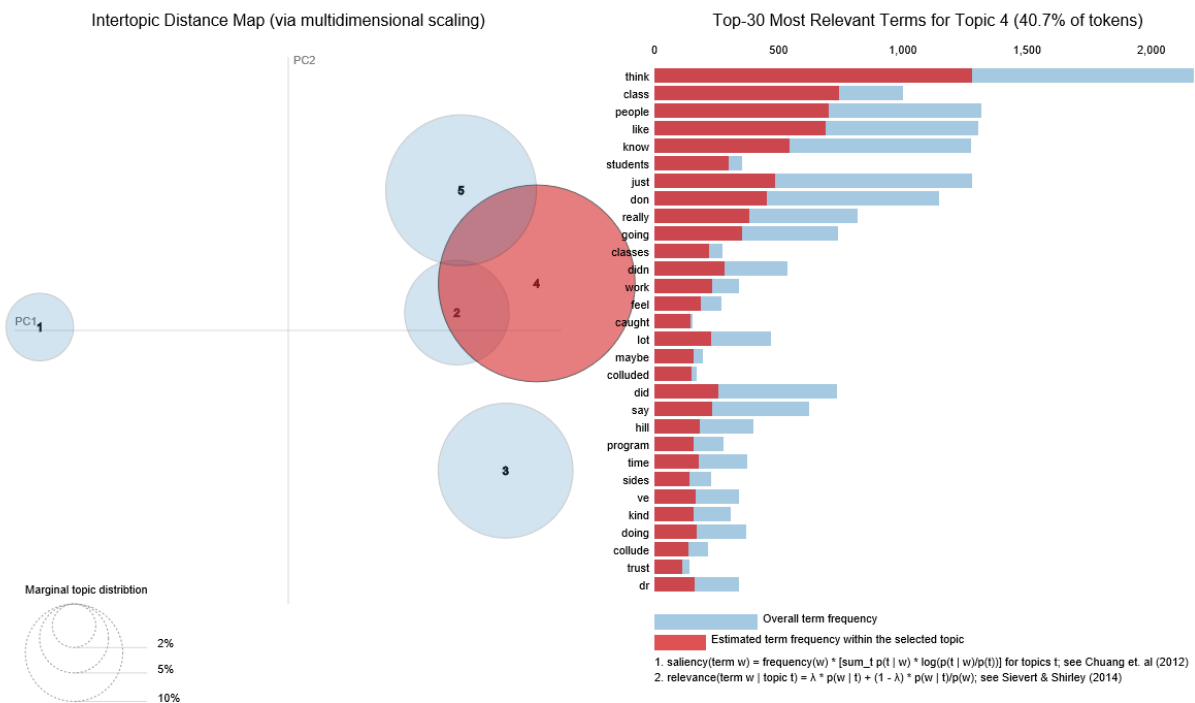
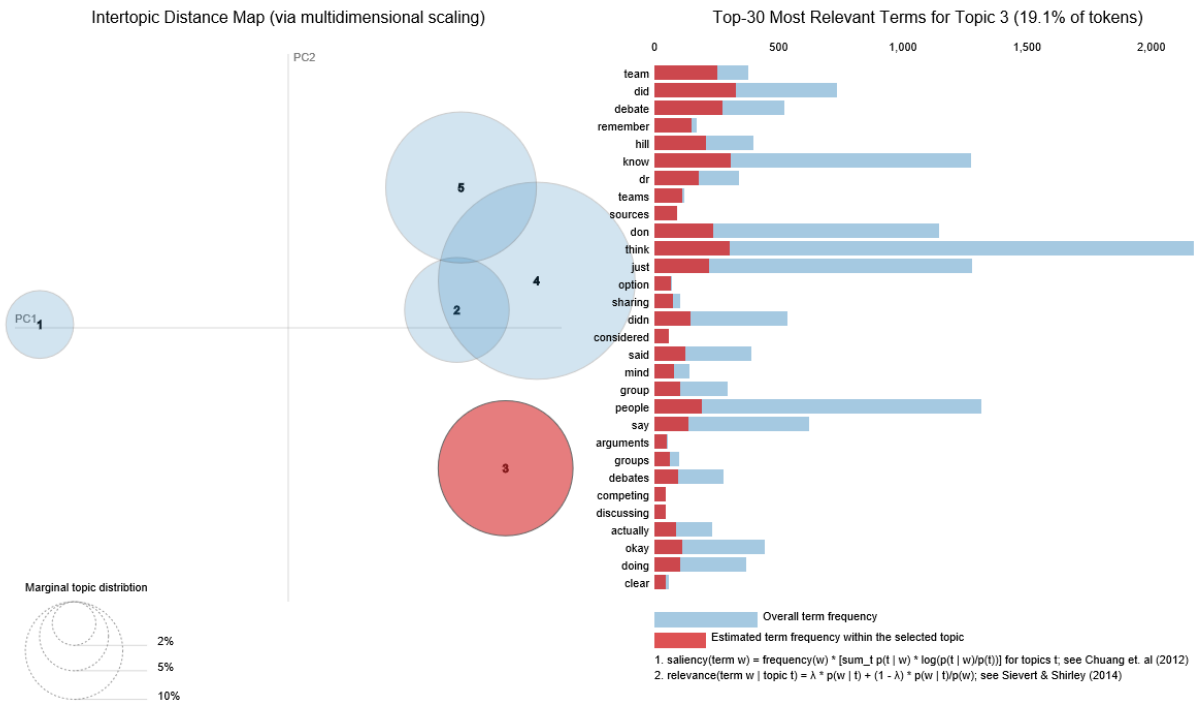
29. Why do you think that you colluded on the debates?

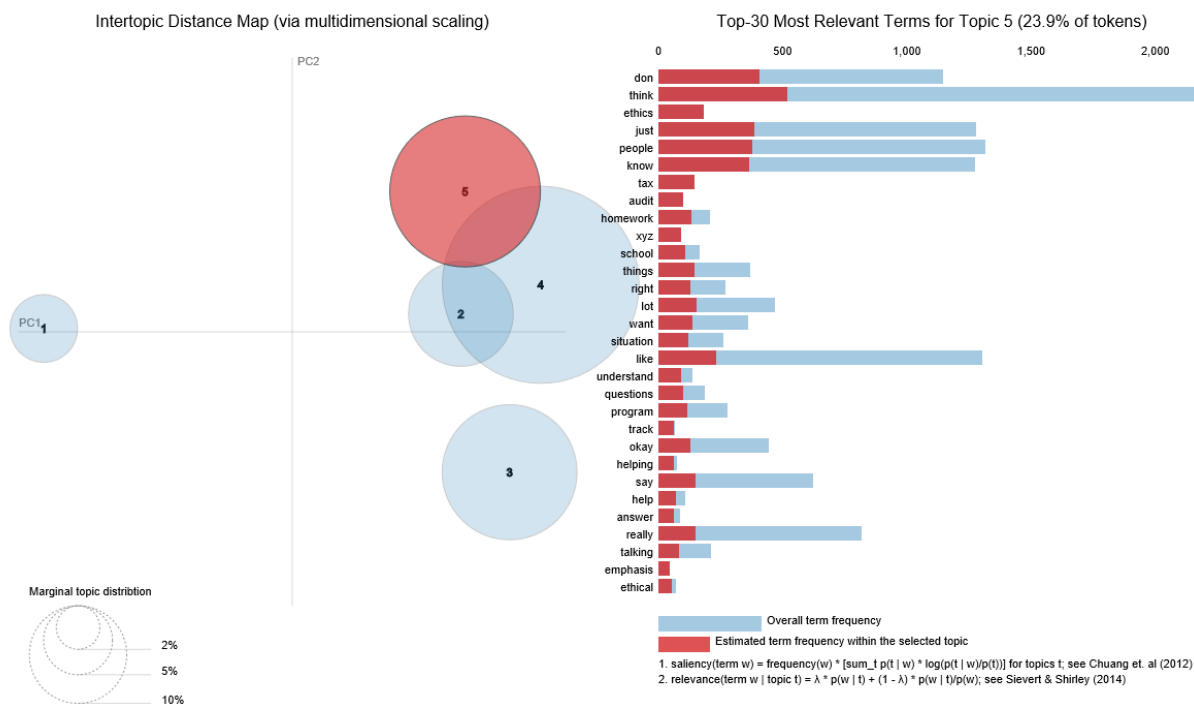
30. What (if anything) do you believe you have learned or will remember about this debate experience?

31. Is there anything else we should know or you would like to say about the debate situation last fall?

APPENDIX B – Maps of 5 topics – most relevant and unique terms per topic







References

- Ariely, D. (2012). *The (Honest) Truth about Dishonesty*, Harper Collins Publishers, New York.
- Ballantine, J. A., Guo, X. & Larres, P. (2018). Can future managers and business executives be influenced to behave more ethically in the workplace? The impact of approaches to learning on business students' cheating behavior. *Journal of Business Ethics*, 149: 245-258.
- BaBever, L. (2017). Dozens of Ohio State students accused of cheating ring that used group-messaging app, *The Washington Post*, November 13.
https://www.washingtonpost.com/news/grade-point/wp/2017/11/13/dozens-of-ohio-state-students-accused-in-cheating-ring-using-group-messaging-app/?utm_term=.e35e9aea7edf
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*. 3: 993-1022.
- Bloodgood, J. M., Turnley, W. H., & Mudrack, P. E. (2010). Ethics instruction and the perceived acceptability of cheating. *Journal of Business Ethics*, 95: 23-37.
- Briggs, K., Workman, J. P., & York, A. S. (2013). Collaborating to cheat: A game theoretic exploration of academic dishonesty in teams. *Academy of Management Learning & Education*, vol. 12, 1, 4-17.

Burrus, R. T., McGoldrick, K., & Schuhmann, P. W. (2007). Self-reports of student cheating: does a definition of cheating matter?, *Research in Economic Education*, winter, pp. 3- 16.

Chen, S. & Wang, Y. (2018). Latent Dirichlet Allocation, Department of Electrical and Computer Engineering, University of California San Diego, <http://acsweb.ucsd.edu/~yuw176/report/lda.pdf>

Cizek, G. J. (1999). *Cheating on Tests: How to Do It, Detect It, and Prevent It* (Lawrence Erlbaum Associates, Mahwah, NJ).

Cronan, T. P., Mullins, J. K., & Douglas, D. E. (2018). Further understanding factors that explain freshman business student's academic integrity intention and behavior: Plagiarism and sharing homework. *Journal of Business Ethics*, 147: 197-220.

D'Souza, K. A. & Siegfeldt, D. V. (2017). A conceptual framework for detecting cheating on online and take-home exams. *Decision Sciences Journal of Innovative Education*, 15, 4: 370-391.

Dyer, T., Lang, M., & Stice-Lawrence, L. (2017). The evolution of 10-K textual disclosure: Evidence from latent Dirichlet allocation, *Journal of Accounting and Economics*, 64, 221-245.

Flynn, R. S. (2003) Confronting academic dishonesty in the accounting classroom: a cultural experience?, *Accounting Education: an international journal*, 12 (4), pp. 437-439.

Genereux, R., & McLeod, B. (1995) Circumstances surrounding cheating: a questionnaire study of college students, *Research in Higher Education*, 36(6), pp. 687-704.

Ismail, S. & Yussof, S. H. (2016). Cheating behavior among accounting students: Som Malaysian evidence. *Accounting Research Journal*, 29, 1: 20-33.

Jeffrey, C., Dilla, W., & Weatherholt, N. (2004) The impact of ethical development and cultural constructs on auditor judgments: A Study of auditors in Taiwan. *Business Ethics Quarterly*, vol. 14, 3, 553-579.

Jones, T. M. (1991). Ethical decision-making by individuals in organizations: An issue contingent model. *Academy of Management Review*, 16, 366-395.

Karlins, M., C. Michaels, & S. Podlogar: (1988). 'An Empirical Investigation of Actual Cheating in a Large Sample of Undergraduates', *Research in Higher Education* 29(4), 359-364.

Kaufmann, J., West, T., Ravenscroft, S., & Shrader, C. B. (2005). Ethical distancing: rationalizing violations of organizational norms, *Business and Professional Ethics Journal*, 24(3), pp. 101-134.

Klein, H. A., Levenburg, N. M, McKendall, M. & Mothersell, W. (2007). Cheating during the college years: How do business school students compare? *Journal of Business Ethics*, 72:197-206.

Lash, M. T. & Zhao, K. (2017). Early predictions of movie success: The who, what, and when of profitability. *Journal of Management Information Systems*, 33, 3: 874-903.

- Lawson, R. A. (2004). Is classroom cheating related to business students' propensity to cheat in the "real world"? *Journal of Business Ethics* 49: 189-199.
- McCabe, D. L., Butterfield, K. D., & Trevino, L. K. (2006). Academic dishonesty in graduate business programs: prevalence, causes, and proposed action. *Academy of Management Learning & Education*, 5(3), pp. 294-305.
- McCabe, D. L. & Trevino, L. K. (1993). Academic dishonesty: honor codes and other contextual influences, *Journal of Higher Education* 64(5), pp. 522-539.
- McCabe, D. L. & Trevino, L. K. (1996). What we know about cheating in college: longitudinal trends and recent developments, *Change*, 28(1): 28-33.
- McCabe, D. L. & Trevino, L. K. (1997). Individual and contextual influences on academic dishonesty: a multi-campus investigation, *Research in Higher Education*, 38(3), pp. 379-396.
- McCabe, D. L. & Trevino, L. K. (2002). Honesty and honor codes, *Academe*, 88(1), pp. 37-41.
- Miceli, M. P., J. B. Dozier, & J. P. Near: (1991). 'Blowing the Whistle on Data Fudging: A Controlled Field Experiment,' *Journal of Applied Social Psychology* 21(4), 271-295.
- Moro, S., Cortez, P., & Rita, P. (2015). Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. *Expert Systems with Applications*, 42: 1314-1324.
- Nowell, C., & D. Laufer. (1997). 'Undergraduate Student Cheating in the Fields of Business and Economics,' *Research in Economic Education* 18(1), 3-12.
- Peterson, D. K. (2002). The relationship between unethical behavior and the dimensions of the ethical climate questionnaire, *Journal of Business Ethics*, 41, pp. 313-326.
- Ponemon, L.(1993). 'Can Ethics be Taught in Accounting?' *Journal of Accounting Education* 11,185-209.
- Premeaux, S. R. (2005). Undergraduate student perceptions regarding cheating: Tier 1 Versus Tier 2 AACSB accredited business Schools, *Journal of Business Ethics*, 62, pp. 407-418.
- Pullen, R., Ortloff, V., Casey, S. & Payne, J. B. (2000). Analysis of academic misconduct using unobtrusive research: a study of discarded cheat sheets, *College Student Journal*, 34(4), pp. 616-625.
- Roig, M. & Ballew, C. (1994). Attitudes toward cheating of self and others by college students and professors, *The Psychological Record*, 44(1), pp. 3-13.
- Scheers, N. J., & C. M. Dayton. (1987). 'Improved Estimation of Academic Cheating Behavior Using the Randomized Response Technique,' *Research in Higher Education* 26, 61-69.
- Schwab, A. & Zhang, Z. (2018). A new methodological frontier in entrepreneurship research: Big data studies. *Entrepreneurship: Theory and Practice*, February, 1-12.

-
- Shrader, C. B., Ravenscroft, S. P., Kaufmann, J. B. & West, T. D. Classroom cheating and student perceptions of ethical climate. *Teaching Ethics*, Vol. 13, No. 1- fall 2012, 105-128.
- Sievert, C., & Shirley, K. E. (2014). LDAvis: A method for visualizing and interpreting topics. Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces-Association for Computational Linguistics. Baltimore, Maryland, June: 63-70.
- Sierra, J. J., & Hyman, M. R. (2008). Ethical antecedents of cheating intentions: evidence of mediation, *Journal of Academic Ethics*, 6(1), pp. 51-66.
- Simha, A. & Cullen J. B. (2012). A Comprehensive Literature Review on Cheating, *International Journal of Cyber Ethics in Education*, 2, (4), 24-44.
- Simkin, M. G. & McLeod, A. (2010). Why do college students cheat? *Journal of Business Ethics*, 94: 441-453
- Spiller, M. S., & D. F. Crown. (1995). Changes over Time in Academic Dishonesty at the Collegiate Level, *Psychological Reports* 76, 763-768.
- Sugimoto, C. R., Li, D., Russell, T. G., Finlay, S. C. & Ding, Y. (2011). The shifting sands of disciplinary development: Analyzing North American library information science dissertations using Latent Dirichlet Allocation. *Journal of the American Society for Information Science and Technology*, 62 (1): 185-204.
- Tirunillai, S. & Tellis, G. J. (2014). Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent Dirichlet allocation. *Journal of Marketing Research*, Vol. LI (August): 463-479.
- West, T., Ravenscroft, S. P., & Shrader, C. (2004). Cheating and moral judgment in the college classroom: a natural experiment, *Journal of Business Ethics*, 54:173-183.
- Woodbine, G. F. & Amirthalingam, V. (2013). Cheat and you lose! Reflections and analysis of accounting student data. *Journal of Academic Ethics*, December, Vol. 11, 4, 311-327.
- Wright, T. A. (2004). When a student blows the whistle (on himself): A personal experience. *Journal of Management Inquiry*, 13, 4, 289-303.
- Zupic, I. & Cater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods*, 18, 3: 429-472.

