Advising of the Students, by the Students, for the Students: The Case of a Student-Owned Peer Advising Community

Alex Duncan
College of Computing
Georgia Institute of Technology
85 5th Street NW, Atlanta, GA 30308
alex.duncan@gatech.edu

David Joyner
College of Computing
Georgia Institute of Technology
85 5th Street NW, Atlanta, GA 30308
david.joyner@gatech.edu

Abstract—Peer advising, where students receive advice from other students about courses, can be difficult to study due to its private, spontaneous nature or to potential response bias. However, peer advising can help create and strengthen student communities, and the information students provide can help improve course quality. We first present a case study on a student-run course review website developed to support an online graduate program in computer science. We discuss the evolution of the website and examine its usefulness as a student-organized community providing peer advising at scale. In our second study, we develop a coding scheme and use it to analyze reviews from the website. Although students provide mostly evaluative information, they also provide advice, context for their reviews, course descriptions, and feedback for the instructional team. This research explores the importance of student-organized communities in higher education and provides useful insights into peer advising at scale.

INTRODUCTION

Academic advising can help students in an academic program make the right choices when it comes to selecting courses and determining an overall path through the program. Traditionally, this advising occurs between an academic adviser and a student. While useful in some contexts, this type of advising may not scale up in larger programs. As a student population grows, traditional individualized advising becomes more time consuming, and advisers may have difficulty keeping up with the expanding institutional knowledge of students. As an academic program grows and changes, the challenge for advisers to keep track of details at both a program level and a course level increases.

Peer advising in the delivery of higher education encompasses activities where, rather than receiving advising by faculty or dedicated advisers, students instead receive advising from other, often more senior, students. As with traditional academic advising, peer advising will include information about course selection and advice about progressing through a degree program, but it may include other relevant information or opinions. Although this advising can be administered as part of official university programs (Barman & Benson, 1981), it also arises in ad hoc and informal mechanisms. One question in online education is specifically whether and how student communities arise (Sun & Rosson, 2017). This community engagement and development is one of the most desirable elements of peer advising: although the content of the advising itself may not be as accurate (Goldberg, 1981), it serves a social role to knit together student communities, which has positive effects on performance and retention (Thomas, Herbert, & Teras, 2014).

This informal peer advising is traditionally difficult to study because it largely occurs in ephemeral conversations in private locations. The content of these conversations may be delicate, especially if students are providing negative feedback about classes they are currently taking, reducing the likelihood that students would retell the same advice in a research setting. Online education can allow more insight because the conversations are inherently self documenting, but they still often occur privately, or on non-anonymous forums where students may self censor to avoid retribution for negative feedback. Even in private forums, separating peer advising from other discussion topics can be difficult due to the fact that these forums are conversational and there is no clear delineation between topics.

In our work, however, we have come across a case of students self-organizing peer advising in a more formal, analyzable structure. As part of a large online graduate program in computer science, students have independently developed a course review website. On this site, students may leave plaintext reviews for the program's courses, annotated by Likert-scale assessments of course quality and rigor and numeric assessments of workload (in hours per week). Similar sites have existed, such as RateMyProfessor, but the site under analysis here has gained wider adoption with nearly 3,000 reviews covering only about 30 courses. We hypothesize this is partially due to the student-run, course-oriented, and program-specific nature of the site, but also due to the scale of the program: at 8,600 students in Spring 2019, there is far larger demand for such a centralized aggregator of past student feedback. In comparison to sites like RateMyProfessor as well, this site appears less susceptible to the negative response bias present on sites in other domains (Yüksel, 2017) and functions more as a peer advising community. This plays a notable role for program scale as well: prior work has noted that efforts toward scaling online programs often focus on individual course delivery rather than program-level infrastructure, including student advising (Joyner, Isbell, Starner, & Goel, 2019). It is worth investigating whether a peer advising community like this one can help fill gaps left in student advising at scale.

Thus, we find this emergent, student-owned peer advising community presents two novel research opportunities: first, the site is a worthwhile case study on student self-organization for a peer advising community, which can help address challenges to supporting learning at scale. Second, the structured nature of the site provides a valuable dataset for investigating what kind of information students share with one another in a peer advising community. In this work, we cover both research directions: first, we perform a case study on the history and current state of the site. Second, we perform a qualitative analysis on the content of the reviews, developing an original coding scheme for review content and then summarizing a significant subset of the reviews according to that coding scheme. The purpose of this research is to: explore and understand informal peer advising communities at scale; discover what information students feel is important to provide to future learners; and gain insights into how the various types of information they provide interact to form the context for a course review. This context can help with understanding how to reliably use student reviews to improve course quality.

RELATED WORK

In this work, we build on the literature investigating peer advising. Previously, this literature has largely focused on university-run peer advising centers (Barman & Benson, 1981). Prior research has even specified that for peer advising to be successful, it must be deliberative and supervised (King, 1993). When properly administered, though, research has found it to be effective (Brown &

Myers, 1975; Murry, 1972), although potentially only as a supplement to traditional advising programs (Goldberg, 1981). Additionally, while much of the foundational research on peer advising was performed several decades ago, these models persist to this day: most modern investigations of peer advising are top-down, deliberately designed, and university-run (Ellis & Gershenson, 2016; Griffin, DiFulvio, & Gerber, 2015).

In online education, however, such deliberately-created programs may not be the only method of peer advising. Significant research has explored the tendency of online communities to emerge without top-down planning (Hersberger, Murray, & Rioux, 2007; Aragon & Davis, 2019). Campbell et al. (2016) investigate this in the context of emergent, distributed, online writing communities and articulate seven key attributes of distributed mentoring that, although developed in the context of a different skill and community, apply to this emergent peer advising community as well; specifically, abundance (the raw volume of responses available), availability (the public nature of the responses), and asynchronicity (the durability of these reviews across time).

Specific to education, research has already been devoted to the behavior of students as they interact in these communities (Almatrafi & Johri, 2017; Bishara et al., 2017; Nistor & Serfain, 2017), although this research has focused on non-emergent communities not specifically dedicated to academic peer advising. There are thus opportunities to: examine peer advising as a function fulfilled by emergent online student communities rather than as a school-created program; to investigate questions regarding how these communities arise and are administered; and to study the content students communicate to one another through these communities. In this analysis, we focus on asynchronous peer advising for individual courses in a massive for-credit graduate degree program on a student-owned course review website. In summarizing review content, we focus on the high-level structure of information students communicate.

BACKGROUND

To understand the context of this analysis, there are two important background elements to describe: first, the novel context of the program itself, and second, the specific challenges noted by advisers in the program.

Program Background

The course review site under analysis in this study was created to support an online graduate program in computer science. It later evolved to also cover online graduate programs from the same university in analytics and cybersecurity. The online computer science program, launched in January 2014, is built for scale: tuition is under $10,000, classes are offered entirely asynchronously, and no residential component is required. As a result, the program has drawn a non-traditional student body: the median age is 35, 85% of students are employed full time, and 80% of students are in the United States. By contrast, the on-campus program has a median age of 23, fewer than 10% of on-campus students are employed full time, and most come to the United States from overseas for the degree (Joyner & Isbell, 2019).

As of Spring 2019, the program has grown to over 8,600 active students with over 11,000 individual course enrollments in the most recent full term. Individual classes have an average completion rate of 80%, slightly lower than the on-campus average of 85%, while the program as a whole has a 70% retention rate (Joyner & Isbell, 2019). The two additional programs now covered by the site supply another 2,000 students; some classes are offered in multiple programs, while others are specific to one program.

Advising Background

In order to understand the potential roles that peer advising might be playing in this program, we analyzed e-mails and other electronic communications with the program's advisers, focusing specifically on challenges they encounter that they had not previously witnessed in on-campus programs.

First, the advisers noted that online students tend to have a weaker understanding of certain foundational university policies and procedures. These details are usually communicated via mandatory synchronous orientations for on-campus students, which is replaced by informational emails and packets online; we speculate that it is more likely for students to overlook details in this asynchronous communication format than in these synchronous mechanisms. We also speculate that this may simply be a product of scale: even if the same fraction of both online and on-campus students have the same questions, the online program admitted 21 times as many students in the most recent school year, dramatically raising the volume of questions.

This leads to a second observation, which is that the adviser-to-student ratio is lower on campus than online. While the online program enrolled 20 times as many students in the most recent school year, it only has three to four times as many advisers. Simply hiring more advisers at the same ratio, however, is not a solution, as advisers have a strong pressure to be consistent in their answers and accommodations. More advisers raises the difficulty of ensuring consistent responses, and students already have a tendency to "shop" their questions around to different advisers hoping for a more preferred answer.

Third, the novel nature of the program and its instructional methods leads to many new questions, such as where students access assignments, what grading policies exist, and what expectations are for academic integrity. Although these details are class-specific, advisers for the on-campus program tend to be familiar enough to know to whom to direct students for these questions. Advisers have noted that many of the questions they receive from online students are of a significantly different kind to those from on-campus students, and there is uncertainty about how to answer them or to whom to direct them. In most cases, they note that peers may have stronger ideas than advisers on many questions.

Fourth, compounding that, advisers find that courses and policies in the online program change significantly faster, and it can be difficult to keep up with all the nuances. Changes are communicated via official channels, but follow-up questions often come in, especially from students who missed the initial communication. Several people have noted that in these instances, it is often wise to rely on the wisdom of the crowd and allow students to answer one another's questions as the students most likely to answer are also those most likely to have kept up with changes.

Many of these observations are actually more heavily addressed through a different peer advising avenue: the rich social media communities that have sprung up around the program. The program's student-run Google+ community totals 9,663; the student-run Facebook community holds 2,426; and the program subreddit counts 6,595 subscribers. A student Slack organization for the program counts over 6,000 members and exceeds 3,500 messages per day; the number of daily messages is even higher during high-activity time, especially around registration. These communities pose an incredible research opportunity as well, but the content of these forums is largely unstructured, with everything from questions to promotional posts to social networking. Thus, in order to start our research on peer advising in this community, we focus first on a tool that provides structured content with clear objectives: a student-driven course review site that functions in large part as a mechanism for helping students share perspectives on coursework with their classmates.

STUDY #1: CASE STUDY ON A STUDENT-RUN COURSE REVIEW SITE

Our first study is a retrospective case study on the development of the student-driven site. This case study involves three data sources: first, our own communications history as we observed its development from afar; second, sources gleaned from internet archive services that show historical snapshots of the site; and third, conversations with the current student administrators of the site to confirm our account and supply additional details. Primarily through the first two data sources, we assembled a technological history of the site's progression and development, which we then augmented with conversations with current and former student administrators to understand the motivation behind some of these changes.

The history of the site comprises three phases, which we dub Version 0, Version 1, and Version 2. These are not terms used by the students but are artifacts of the narrative we assembled. Traffic to all three versions of the tool was driven by links added to the descriptions of students' primary social groups (Google+, Reddit, Facebook, HipChat, and later Slack).

Version 0: A Collaborative Spreadsheet

The original version of what would become the student-run course review site was a collaborative workbook created in Google Sheets in July 2014. This sheet was created by a student for personal use shortly after the program's first semester, but it quickly became a tool for other students to use, reflecting an early interest in more structured and shared peer advising. Students initially created one sheet per class, and each sheet had columns for difficulty, workload, general comments, and reviewer name (explicitly labeled as optional). Starting the following term, a column was added to reflect what term the class was taken; this column helped capture course changes over time. No validation was present to enforce how students assess workload and difficulty, but students nearly universally rated difficulty on a 7-point Likert scale: Very Easy, Easy, Somewhat Easy, Average, Somewhat Hard, Hard, and Very Hard. There were exceptions: students sometimes noted "See Comments" for difficulty, showing they were not comfortable assigning a simple label. In other rare places, students added an extra annotation, such as "Artificially Hard", "As hard as you want", "Topics: Easy; Workload: Difficult", and "Not difficult; boring and time-consuming". There is some social conformity here: for example, the last two labels each appeared for only one course, but they appeared multiple times for that course.

For workload, students generally specified ranges like "10 to 20" or "15+" and occasionally annotated these with additional details, such as "up to 40 during project weeks".

Prior to being deprecated in favor of what we refer to as Version 1, Version 0 received 500 reviews for 16 classes. The sheet in this form made the abundant content persistently and publicly available, embodying three of the principles of distributed mentoring noted by Campbell et al. (2016). The last reviews present in Version 0 are stamped as covering classes in Fall 2015; thus, we infer (though our conversations with the administrators were unable to confirm this) that Version 1 was built in late Fall 2015 and began receiving reviews in Spring 2016.

Version 1: The Original Website

The original dedicated course review site was built as a web application on the popular Heroku platform. The site was built by one student to create what he saw to be a more useful interface for inputting and navigating course review data. Several changes were implemented in Version 1. First, students were asked to rate course quality in addition to difficulty and workload. Second, "difficulty" was reframed as "rigor". Third, validation ensured that rigor and quality were enforced on 5-point Likert scales, while workload accepted only a single integer as input. Prior reviews from the spreadsheet were migrated into the new site but were not annotated with numeric assessments for rigor and quality (even if labels had been assigned in the previous spreadsheet). The prior spreadsheet was then labeled as deprecated, and a message and link were placed on the first sheet.

Additionally, a dashboard was created listing all courses with the number of reviews and their average rated quality, rigor, and workload, by which students could sort and filter the courses. A page was added listing reviews in reverse chronological order, allowing students to keep up with the most recent reports. Interestingly, this feature was implemented at the request of an instructor in the program, and program staff then used this page frequently to keep up with any changes or emerging issues in the program. This was the first documented occurrence of the school significantly using the student-run website. It followed similar trends on other sites, where program staff began answering student questions in student-owned social media groups.

One major challenge emerged with Version 1: the program was growing, and there were small windows of time when the content of the site was most relevant to students. As a result, the site reliably went down during registration when too many students wanted to access it at once. Because of the structure of Heroku's hosting, once the allotted bandwidth was exhausted the site would remain down for the rest of the calendar month. This presented a significant challenge, and students compensated by keeping local copies of the course review pages.

Version 2: The Second Website

During one such period of downtime in Fall 2017, another student in the program wrote to the original author (both of whom participated in our conversations for this study) and offered to help with the site. The original author responded by offering to transfer ownership to the new student, who accepted the offer. This new student owner rewrote the code for the site and imported all prior reviews. To address the bandwidth issues, he improved the data efficiency of the site and required students to log in to see existing reviews to lower the number of requests. He also connected the site with the university's historical grade database and added grade distributions and completion percentages for the past several semesters. The new author added functionality to filter courses by program or specialization and to filter reviews by semester, difficulty, or quality. These improvements allowed students to answer questions like: how was the course last term? What do people who dislike the course have to say? What do people who like the course recommend we do to prepare?

Figures 1 and 2 show screenshots of the website in its current form. This version of the website has run from Fall 2017 to the time of writing and has gathered 2,903 reviews. It is unclear how many of these reviews were originally authored for Version 1 and how many were provided only once Version 2 arrived. The university launched a second online Master's program in Fall 2017 and, because the site automatically drew all online courses into its database, this new program became available for review on the site. Those courses solely associated with the new program account for 126 reviews, although other courses are shared between the programs and may have reviews from students in each.

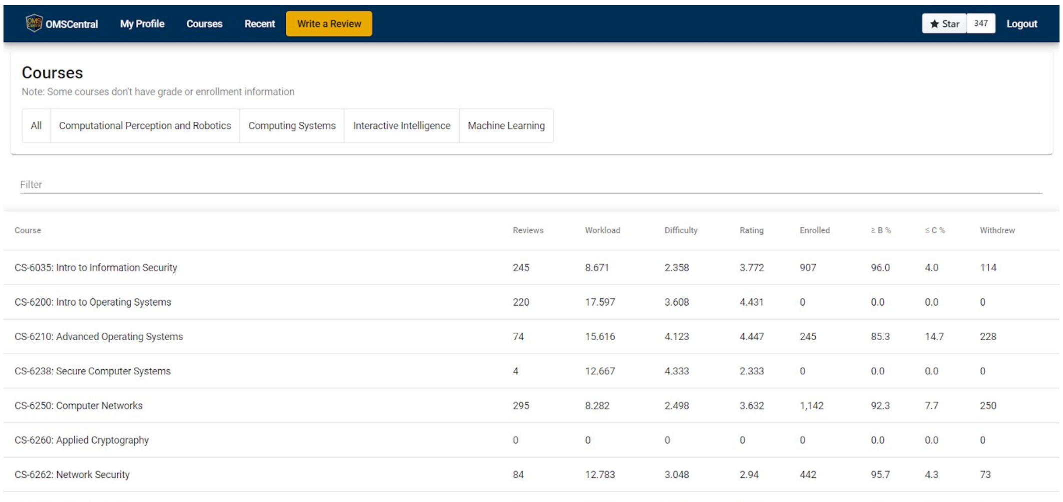

Figure 1. A list of the courses on the course review website, showing the search and filter options.

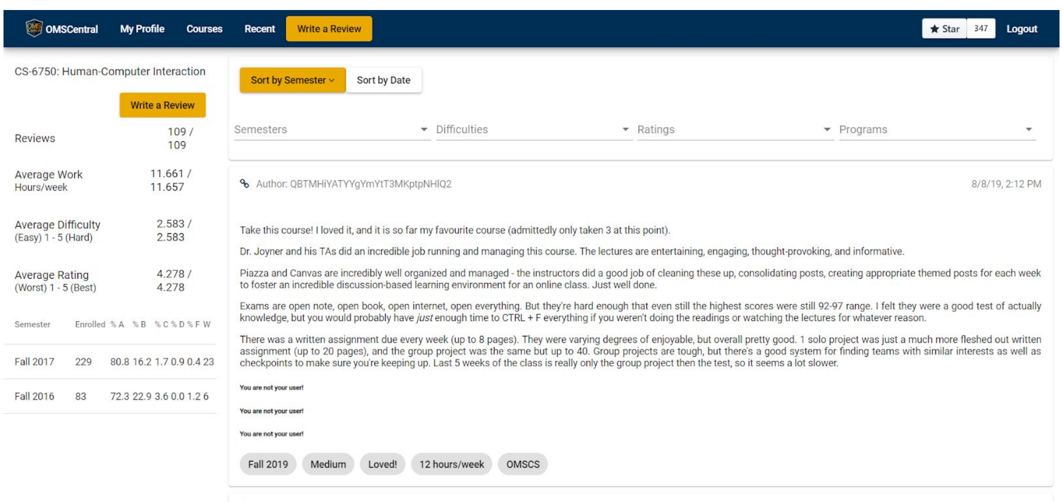

Figure 2. The view of a single course on the course review website, showing course summary information and a list of all reviews that can be sorted and filtered.

Discussion

The history of the website has several noteworthy events. First, the creation of a shared spreadsheet in the first place; second, the implementation of a standalone site containing the data from the first spreadsheet; and third, the creation of a new site to address the issues encountered with the previous site. Each major development was sparked by one person but was adopted by the whole community, suggesting the site may be a potential alternative or addition to traditional, individualized advising, which is difficult to scale to large student populations. It is also noteworthy that the tool was able to reach such a large audience through student-driven communities alone. This is a unique factor in an online community, where the entire community can interact in a centralized space, magnifying individual voices and directing attention to resources such as this site. Finally, it is notable that while many such projects devolve into disarray when the original architect departs, this site has undergone two successful transitions. In one, the transition occurred because of the creation of a superior tool; in the other, the transition occurred because of a proactive move by the original architect to transfer ownership. These types of transitions appear necessary to sustain these community-led initiatives for the long term.

STUDY #2: UNDERSTANDING THE CONTENT OF PEER ADVISING

Our second study examines the content of peer advising. As noted above, this site presents a unique opportunity to understand the information and advice students give one another: it is anonymous, allowing students to reflect more openly and honestly, but it is also permanent and public, allowing it to be evaluated rigorously. In our analysis, we focused on the plaintext content: the numeric ratings will provide an additional variable for future analysis, but for this analysis we were interested in summarizing the review content. To do this, we developed a coding scheme, performed intra-rater and inter-rater reliability checks on the coding scheme, and applied the coding scheme to the full dataset. Here, we present those results and summarize the trends we observed.

Methodology

Our methodology included three tasks: developing the coding scheme, validating the coding scheme, and coding a significant chunk of the dataset. These tasks were performed on a partial export of the review database including 1,373 reviews. Reviews from the original spreadsheet were excluded due to incomplete and inconsistent data. This export is smaller than the number of reviews noted above due to timing: the analysis took place over several months, while the numbers above reflect the number of reviews at time of writing.

Developing the Coding Scheme

The coding scheme was developed in two phases. First, the reviews were exported and randomized, and the first coder pulled out an initial dataset of 50 reviews. Based on these 50 reviews, the first coder identified four common categories of content in the reviews: Advice, where a reviewer gives direct and actionable advice to the reader; Course Description, where the reviewer gives objective and useful information about the course, such as whether it includes proctored tests or group projects; Feedback, where the reviewer suggests alterations that the teaching team ought to make to the course; and Other/Low-Relevance, which often included information not pertinent to the review contents itself.

The first coder also observed that reviews typically play multiple roles, and thus it was decided that future coding should instead break the reviews into individual sentences for coding. Then, we randomly selected 50 more reviews and divided them into 480 sentences. These sentences were given to a second coder who identified two additional categories of statements: Review Context, wherein readers provide information useful to the reader to individualize the review contents, such as noting their own technical competence with the class's programming languages; and Evaluation, a variation of Feedback where the reader does not make suggestions for the course, but rather comments on its overall merits and value. These two categories were added to the coding scheme, and we randomly selected another 25 reviews (comprising 214 sentences). No additional categories were found in this set of 25 reviews, so we deemed the coding scheme appropriate for coding reviews for the dataset. The 75 reviews divided into sentences and used to develop the coding scheme were excluded from future coding, although reviews from the initial set of 50 may have reappeared due to details of how we handled the sampling, random selection, and removal of previously coded reviews. The codes were defined as shown in Table 1.

In developing this coding scheme, assumptions were made regarding the intention of certain statements; for example, the statement "I have no formal CS

Table 1. Coding scheme definitions and examples.

Definition	Examples
Advice - Recommendations involving prerequisite knowledge, courses to take before or in conjunction with a course, or the best way to progress through a course were included in this category. Also included were warnings or reassurances about taking a course and information about future offerings of a course. Advice is particularly characterized by being targeted at the reader.	*"I would recommend this course if you have no prior software engineering background, though I would skip it if you have had an undergrad SE class or SE work experience."* *"If you only know a scripting language such as JavaScript, R, or Python, you might want to hit the first project hard, right away, to climb that learning curve."* *"Don't be afraid if your agent can't solve all the problems, getting 50 75% of the problems right is usually good enough to get an A."*
Review Context - This category included statements about the reviewers themselves, such as their coding or professional experience. Facts that were specific to one semester of a course (and therefore possibly not generalizable across all semesters) were also included, along with non advice, non evaluative statements about how the reviewer or other students progressed through a course.	*"Because I took this in the summer, the workload was much higher than you will see in other reviews."* *"I have no formal CS background."* *"While it did suffer some "new class" organization issues (it is only the second semester offered), it looks like the professor and TA's are really working to clean up these issues."*
Course Description - This category contained statements that provide objective information about a course, such as the number of projects or average final grade in a course. Any factual statements about a specific semester of a course that are likely generalizable across many other semesters are also included.	*"The class has 3 assignments, 3 projects, and a midterm and final."* *"All of the coding is done in Java, using the IntelliJ and Android IDEs."* *"If you stay above the mean, you get an A."*
Evaluation - Statements in this category were subjective and related to the reviewer's opinion of a course, often involving likes or dislikes. Statements about a reviewer's dislikes were only included in this category if they weren't actionable; otherwise, they were grouped into the Feedback category.	*"The concepts presented in the lectures aren't too difficult to wrap your mind around, and I found many of them very interesting."* *"This class can get a bit boring."*
Feedback - Actionable statements regarding aspects of a course that a reviewer disliked or wanted changed were grouped into this category. These statements were broadly applicable and not specific to one student. Although most of these statements were usually recommendations for course changes, some of them were about aspects of a course that were beneficial and should not be changed.	*"Since the projects are fairly trivial, it seems silly to spend so much time on design and discussion, so hopefully in the future they use projects that are a bit more complex, or maybe having to deal with changing requirements or incomplete requirements."* *"I'd suggest slowing down some of the more critical lectures or providing more examples."*
Other - Statements that didn't fall into any of the other 5 categories were grouped into this category. These were often post semester outcomes, musings about education as a whole, or fragments that only make sense in the context of previous sentences.	*"Also, because of the practical side of this class, I was able to get a job as a Junior Data Scientist!!"* *"In real life requirements tend to be vague and it takes a good analyst to fill in the blanks."*

background" is regarded as a statement of review context, intended to communicate to the reader that the student's experience may vary for someone with a formal computer science background. Rather than context, however, this statement could serve other functional roles: to justify one's own failure, to describe the class's target demographics, or to emphasize a subsequent statement. We acknowledge these weaknesses in inferring too much about writers' intentions, but we also observe that regardless of intention, these statements play certain functional roles in the mind of the reader as well which may diverge from the intentions of the author.

Coding the Data Set

52% of the dataset (excluding the reviews used to create the coding scheme) was coded; these reviews were selected randomly and thus reach the point of saturation and mirror similar samples from the field (Caine, 2016). These reviews covered 36 courses across two programs. Again, the reviews were coded at the sentence level, which provided more granular information than review-level coding would. This led to 6,746 sentences being coded for the dataset.

Assessing Intra- and Inter-Rater Reliability

We took two approaches to assessing reliability: intra-rater reliability and inter-rater reliability. 74 reviews (approximately 10%) comprising 748 sentences were randomly selected from the coded dataset and re-coded by the original rater approximately a week after coding the original dataset. To assess intra-rater reliability, the original codes were compared to the new codes, and the Kappa statistic was calculated to be 0.757 (Viera & Garrett, 2005), indicating substantial strength of agreement (Landis & Koch, 1977). To assess inter-rater reliability, a second coder used the same coding scheme to code the same 74 reviews. The codes applied by the second coder were compared to the original codes applied by the first coder, and the Kappa statistic was calculated to be 0.578, indicating moderate strength of agreement. These values were deemed sufficient to draw notable descriptive conclusions from the larger 6,746-sentence dataset.

In performing these reliability checks, we observed two main challenges in coding reviews reliably. First, while our definitions distinguish between Evaluation and Feedback, in practice many sentences can be seen as examples of both categories. Feedback is rarely purely objective, and even subjective Evaluation can be used to inform course revision and improvement. For example, one sentence read, “The lectures are good, but the projects (especially project 1) can feel a little irrelevant to what's being covered in the lectures.” While the student is not directly suggesting a change as clearly as “I’d suggest slowing down some of the more critical lectures”, the feedback nonetheless could be used by an instructor to improve the course. To check this, we combined Feedback and Evaluation into one category and recalculated the inter-rater reliability Kappa statistic to be 0.650, indicating that differentiating Feedback and Evaluation was a major challenge in coding reliably.

Second, we found that while coding at the sentence level allows an objective segmentation of reviews, individual sentences often contain multiple types of information. For example, one sentence from one review stated, “I agree with most of the other comments mentioned: workload is do-able (plan at least 12-15 hours a week, as the projects take far longer than you will anticipate), there are roughly 9 hours of video total, and the instructor is very active and very reasonable.” Although that is one sentence, it contains Evaluation (“workload is do-able”), Advice (“plan at least 12-15 hours”), Description (“there is roughly 9 hours of video total”), and Feedback (“the instructor is very active and very reasonable”). Disagreements between raters often arose from raters focusing on different phrases in the same sentence, while still agreeing that multiple segments were present.

Results

We summarized the coded dataset in three ways: the total frequency of reviewers’ usage of each code; the frequency of each code in an average review; and the total frequency of reviews featuring each code at least once. Each of these summaries provides slightly different insights.

Figure 3 summarizes the total number of usages of each code (that is, the total number of sentences assigned each code). The largest chunk of statements is coded as Evaluation, confirming that students focus on providing personal impressions of the course. Objective course description information is the next most-common, suggesting that students prioritize objective criteria (such as whether the class has group projects) when deciding how best to describe and discuss a course. Although no formal message has been communicated that program staff read the site, there is also a significant amount of school-targeted feedback.

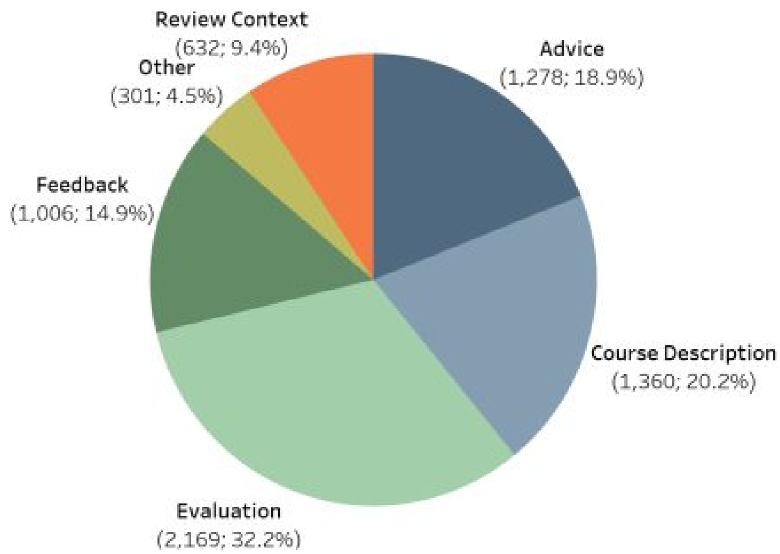

Figure 3. Content of reviews based on total code usage.

Figure 3, however, may be skewed by lengthy reviews and may not show the makeup of a "typical" review. Figure 4 summarizes the average makeup of each review. For this summary, we processed each review based on the number of sentences with each code and then averaged the reviews. This way, the makeup of longer reviews is weighted the same as the makeup of shorter ones. In this summary, we see each category drop one or two percent except for Evaluation, which grows. This suggests that each review tends to be 38% Evaluation, although there are individual reviews that focus more heavily on other components.

Finally, Figure 5 summarizes what fraction of reviews have at least one example of each code, addressing the potential confound that there may be more to say about certain factors than others. Almost all reviews (95.7%) contain some evaluation. Interestingly, the other codes are all present in a much greater number of reviews than their relative percentage of all codes. For example, while Advice is the third-most common overall code, it appears in the second-most reviews: only 19% of sentences are coded as Advice, but 70% of reviews have a sentence coded as advice. This suggests that while students may have more to say about their subjective evaluations, they prioritize directed advice as well.

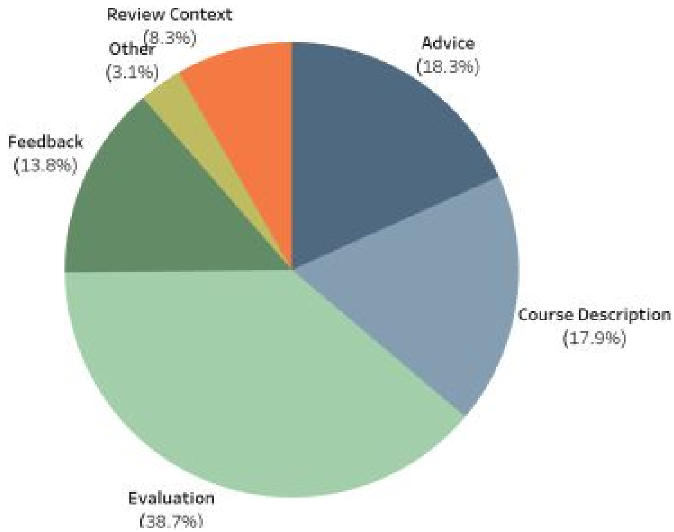

Figure 4. Content of reviews based on average review makeup.

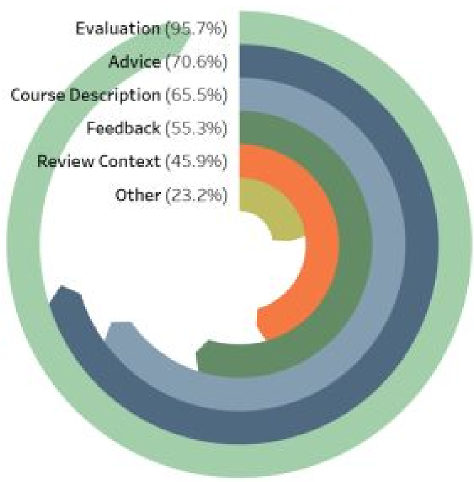

Figure 5. Content of reviews based on percentages of reviews with at least one example of each code.

Discussion

Statements coded as Evaluation are the most common types of statements and appear in nearly every review, suggesting that students are most interested in sharing their opinions of the course. However, this does not necessarily imply that this kind of information is what readers seek in these reviews. Additionally, statements coded as Advice, Course Description, and Feedback each appear in over half of all reviews, with Advice statements appearing in over 70% of reviews. This suggests that, although these statements do not constitute a large part of the average review, students still find this kind of information important to include in reviews most of the time. The information is perhaps easily conveyed in a few sentences, whereas evaluative information, containing more emotion and being more difficult to describe, requires more sentences. Nevertheless, students' tendency to provide advice in most reviews strengthens the impression of the website as a peer advising community, rather than simply a feedback repository.

No category represents a majority of the coded statements, suggesting an interplay between the types of statements and a desire among students to understand the perspective of a reviewer to gauge the validity or applicability of a review. For example, it may not be enough for a student to hear from many reviewers that she should not take a course; the rationale for why she should not is significant. If the reasoning in those reviews is that there is too much reading in the course, but this student enjoys reading, she may place less emphasis on the advice of the reviewers. On the other hand, if the reasoning is that the course is time-consuming, and time commitment is an important factor to the student, she may weigh the reviewers' advice more heavily.

Course descriptions make up 20.2% of the coded statements – the second-highest of the code categories – and appear in 65.5% of reviews, suggesting this information is not readily available or, if it is, changes frequently enough to potentially be unreliable. It may be unclear when publicly available information such as a course syllabus was last updated; on the other hand, course reviews are timestamped and tied to a specific semester. Students seem to find it valuable to provide this information with the added temporal context.

Most reviews contain statements providing feedback to the instructional team of a course, implying that reviewers suspect or hope that the instructional team will read these reviews. However, the context of this feedback relates to the difficulty of determining whether a statement qualifies as Feedback or Evaluation. It is possible that this feedback was provided as part of an evaluation, or possibly as implicit advice. There are many examples of these types of ambiguous statements, such as: "One thing that makes it hard is the pace the assignments are released / delivered throughout the course, every week." The student may intend this as feedback for the instructional team on how to improve the course experience for all students; or, he or she may be evaluating the pace of the course as unnecessarily quick based on his or her own standards; or he or she may want it to serve as a warning to other students that the course is fast-paced without necessarily providing any indication regarding whether he or she liked or disliked the pace. So, care should be taken to understand the nuance in these reviews if they are being used for suggestions on how to improve courses.

FINDINGS IN CONTEXT

With these models in mind, we close this study by attempting to put these findings back into the program context to understand the function that this site is serving. There are two relevant ways in which we find this site integrated into the program context: one, as part of the program's official communications, and two, as one piece of the broader peer advising community surrounding the student body.

Official Usage

Through this study, we have noted that this site documents useful information about courses in the program to share with prospective students and future classmates. We note also, however, that a significant portion of this content comprises facts and perspectives that the program or courses would not likely share officially. Official communication has a tendency to be perceived as binding by students, and it is not uncommon to field student complaints that prerequisites were not adequately delineated or that they were unable to succeed despite meeting the expectations set forth for the class. Leveraging the student community, however, provides an avenue to officially support students without making unfulfillable promises.

Toward this end, the program has started incorporating the content of this peer advising site into official course documentation regarding the individual courses. Other sites already existed to provide high-level descriptions of course topics, expectations, and prerequisites, but the program itself would not aim to assign estimates on workload, time required, assessment style, and other more detailed aspects of the student experience. By referring students to this student-run community, however, the program supports student information-seeking without binding itself to enforcing promises like the amount of work required per class. Notably, these kinds of tools have existed in the past, but they largely leveraged public grade data and official institute course review data and were not widely used; the student-driven nature of this tool appears to drive more attention and a perceived greater credibility due to the absence of the university as a filter on the site's contents. In this context, the potential benefits of the tool are two-fold: a reduction in the burden on university advisers to provide useful and accurate course-specific information to a large student population, and a more direct avenue for students to find answers to their questions.

In The Broader Peer Advising Community

As referenced above, this tool is one piece of a broader peer advising community that exists in this program. It receives and organizes students' perspectives on individual courses, merging them with institute data on grade distributions. This, however, is only one component of peer advising. To close this analysis, we examine the ways in which the tool is referenced in the program's other peer advising components.

To support this, we observed a few ways in which the tool is referenced across other elements of the community. First and most obviously, the tool is referenced often when students ask questions that are directly answered by reviews in the tool. For example, one student asked, "What classes would you recomend [sic] for new student, Machine Learning focus, Spring 2019 semester?" A classmate replied, "Read [the review site] for course reviews which include estimates of course difficulty and (time) effort required."

Second, the tool is frequently referenced in responses to posts from self-confessed new or incoming students, even if the nature of their questions is unrelated to the functions that the site serves. For example, one student recently asked an unrelated question regarding whether they can defer admission, and a classmate replied with, "Please see the course offerings, see [the review site] for guidance about the difficulty and time commitment for courses, and pick several (5-6) courses that you would be willing to take." This response was not relevant to the student's question, but the respondent considered it relevant for any new student. Our perception here is that these are instances where the respondent considers the site valuable enough that it is worth promoting to new students at any opportunity regardless of its immediate relevance.

Third, the tool is often used as evidence for a stated opinion on a class. For example, one recent thread saw a student asking classmates for what class might be an "easy A" in order to raise their GPA. In one response arguing two courses in particular were harder than others perceived, a respondent wrote, "Only about 25 to 30% of students earned an A in [one class], while about 1/3 earned an A in [the other class]", along with a link to the review site. In these contexts, students appear to use the tool to argue that certain minority opinions may not reflect the views of the majority of students. Relatedly, sometimes students cite the review site as evidence of a more factual statement. For example, a student may ask if a certain class has group projects, and a classmate without personal familiarity with the class may nonetheless respond, “According to [the review site], it does.”

These external references suggest that, more than functioning as a peer advising community on its own, this course review site instead is the collective memory of a broader peer advising community spread across multiple media, including subreddits, Slack and Discord organizations, and social media groups. Borrowing the notion of distributed cognition from Hollan, Hutchins, and Kirsh (2000), the course review site is in many ways the distributed memory of the program’s student community.

CONCLUSION

In this work, we performed two studies: first, a case study on the emergence of a student-run course review site that we argue functions as an emergent, distributed peer advising system; and second, a qualitative analysis of the content of student reviews entered into the system to gain insight into the types of advice students give to one another through this community.

Contributions

This work provides three major contributions to the existing literature on peer advising. First, it provides a case study on the emergence of a student-run tool for course-based peer advising. In the process of that development, students addressed several issues with no top-down edict for features or uptime. The persistence and adoption of the system despite its entirely student-run nature and its series of transitions between owners provides an interesting example of student-run peer advising in an online program and suggests a viable method of providing informal advising at a large scale.

Second, we contribute a coding scheme for interpreting student course reviews, and the results of applying this coding scheme to one such dataset of reviews. We find a significant prevalence of subjective impressions (Evaluation and Feedback) that appear to serve as anchors for community-building rather than targeted and actionable advising, which contribute to a cumulative, community-generated collective perspective. We also find a prevalence of reader-targeted subjective impressions in the form of Advice, as well as more objective declarations intended to share with readers information necessary to make a personalized informed decision (Course Description and Review Context).

Third, we contribute an application of the idea of distributed communities of mentorship to this peer advising community. We observe through this website three of the attributes of distributed mentoring (Campbell et al., 2016) in action: the reviews exist in abundance in a publicly available and asynchronous structure. When merged with the role that this website plays in the broader peer advising community (spanning subreddits, Facebook groups, Slack and Discord organizations, and other student-run community efforts), we see also other elements of distributed mentoring in action; accretion and acceleration in particular occur in conversations spawned by or in parallel to reviews aggregated on the site, capturing a diversity of viewpoints. Further work ought to adopt and apply the theory of distributed mentoring to the whole of the program's student community.

Limitations

We would not claim that the distribution of types of content in course reviews would generalize to other levels or programs. This dataset comes from a graduate-level computer science program replete with technical prerequisites for taking certain classes. These prerequisites may not be present in non-technical majors, and we might see significantly different trends in advice-giving in other content areas. We also note that the reviews used in this study come from a distinctive population: mainly graduate computer science students, many of whom are employed or have professional experience. At younger levels, we may see course reviews focused more on workload and requirements and less on course quality. We do not, however, believe that the codes created here would be inaccurate for these new areas; in fact, we hypothesize that this coding scheme may provide a baseline through which we may compare programs at different levels and in different subject areas.

Future Work

Once tested and validated at other levels and in other subject areas, our coding scheme may be useful for comparing student experiences (or at least how students report their experiences) across different areas. It may even be applied in this way to the existing dataset: are there different patterns in the kinds of content students include in reviews for different courses?

Secondly, we hypothesize that Feedback and Evaluation are difficult to differentiate because they are subcategories of a higher-order category, which we tentatively call Assessment, that includes non-objective appraisal of the course. Feedback and Evaluation differentiate whether that assessment is instructor-targeted or student-targeted. We similarly believe there are subcategories for the other codes. Advice, for example, may be differentiated into advice on: prerequisite knowledge; how to approach the course after enrollment; or the expected meaning of certain grades or behaviors. Developing the coding scheme at finer levels of granularity would provide more digestible, accessible advice to student readers and more actionable, well-defined suggestions to instructors. This finer granularity may also be used to inform the creation of a formative feedback mechanism to provide suggestions to instructors more quickly, in the interest of more immediate course improvements.

Third, correlating Likert scale ratings (of course quality, rigor, and workload) on the website with the reviews may provide a fuller picture of students' impressions of a course. Additionally, the university provides students with the opportunity to complete end-of-course instructor surveys, and it may be interesting to understand the differences between student responses to these surveys and reviews on the student-run website. Related research has found interesting trends in this regard, specifically that private instructor-targeted reviews tend to be more positive while public, classmate-targeted reviews tend to be more negative (Newman & Joyner, 2018).

Fourth, because reviews are coded at the sentence level, it is possible to distill reviews down into their underlying structure. We may interestingly find common patterns in the structure of reviews, such as Context > Description > Evaluation > Advice. These patterns would provide a fuller picture of the content of peer advising interactions and describe how some components of a review set up or contextualize other components.

Fifth, our analysis is largely limited to what students feel is worth sharing – we are unable to draw any conclusions about what students want to read. Further analysis may ask readers to evaluate review usefulness. The results may allow for comparisons between the peer advising content students choose to share and what content they find useful. Findings may inform the design of similar review sites that encourage reviewers to provide the kinds of information that readers want to receive.

Finally, these directions are all focused specifically on the analysis that can be performed on these course reviews. As noted, however, these course reviews exist in a much richer peer advising community, the scope of which is too broad for this analysis. Future work will also expand the data under consideration in evaluating online emergent peer advising to include these communities. What kinds of questions do they ask each other? Who answers these questions? Are there a small number of power users or influencers, or are answers relatively well-distributed among many different respondents? How do these response patterns change over time? These data will provide an even fuller picture of peer advising at scale.

REFERENCES

1. Almatrafi, O., & Johri, A. (2017). Showing and Telling: Response Dynamics in an Online Community of Makers. Philadelphia, PA: International Society of the Learning Sciences.
2. Aragon, C. & Davis, K. (2019). *Writers in the Secret Garden: Fanfiction, Youth, and New Forms of Mentoring*. Boston, MA: MIT Press.
3. Barman, C. R., & Benson, P. A. (1981). Peer advising: A working model. *NACADA Journal, 1*(2), 33-40.
4. Bishara, S., Baltes, J., Hamid, T., Li, T., Nacu, D. C., Martin, C., Gemmell, J., MacArthur, C., Raicu, D., & Pinkard, N. (2017). Revealing Interaction Patterns Among Youth in an Online Social Learning Network Using Markov Chain Principles. Philadelphia, PA: International Society of the Learning Sciences.
5. Brown, C. R., & Myers, R. (1975). Student vs. faculty curriculum advising. *Journal of College Student Personnel*.
6. Caine, K. (2016, May). Local standards for sample size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 981-992). ACM.
7. Campbell, J., Aragon, C., Davis, K., Evans, S., Evans, A., & Randall, D. (2016, February). Thousands of positive reviews: Distributed mentoring in online fan communities. In *Proceedings of the 19th ACM Conference on*

Computer-Supported Cooperative Work & Social Computing (pp. 691-704). ACM.

8. Ellis, J. R., & Gershenson, S. (2016). *LATE for the Meeting: Gender, Peer Advising, and College Success* (No. 9956). Institute for the Study of Labor (IZA).
9. Goldberg, L. G. (1981). Peer advising: A supplement to, but not a substitute for, faculty advising. *NACADA Journal, 1*(2), 41-43.
10. Griffin, M., DiFulvio, G. T., & Gerber, D. S. (2015). Developing leaders: implementation of a peer advising program for a public health sciences undergraduate program. *Frontiers in public health*, 2, 288.
11. Hersberger, J. A., Murray, A. L., & Rioux, K. S. (2007). Examining information exchange and virtual communities: an emergent framework. *Online Information Review, 31*(2), 135-147.
12. Hollan, J., Hutchins, E., and Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM TOCHI 7*, 2: 174-196.
13. Joyner, D. A. & Isbell, C. (2019). Master's at Scale: Five Years in a Scalable Online Graduate Degree. In *Proceedings of the Sixth Annual ACM Conference on Learning at Scale*. Chicago, Illinois, USA.
14. Joyner, D., Isbell, C., Starner, T., & Goel, A. (2019). Five Years of Graduate CS Education Online and at Scale. In Proceedings of the ACM Global Computing Education Conference (CompEd). Chengdu, China. ACM Press.
15. King, M. C. (1993). Advising models and delivery systems. *New directions for community colleges, 1993*(82), 47-54.
16. Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159-174.
17. Murry, J. P. (1972). The Comparative Effectiveness of Student-to-Student and Faculty Advising Programs. *Journal of College Student Personnel, 13*(6), 562-566.
18. Newman, H., & Joyner, D. (2018, June). Sentiment Analysis of Student Evaluations of Teaching. In *International Conference on Artificial Intelligence in Education* (pp. 246-250). Springer, Cham.
19. Nistor, N., & Serafin, Y. (2017). Newcomer Integration Strategies in Blogger Online Knowledge Building Communities: A Dialog Analysis. Philadelphia, PA: International Society of the Learning Sciences.

20. Sun, N., & Rosson, M. B. (2017). Finding the Community in Online Education: It's in the Instructors' Eyes. Philadelphia, PA: International Society of the Learning Sciences.
21. Thomas, L., Herbert, J., & Teras, M. (2014). A sense of belonging to enhance participation, success and retention in online programs.
22. Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Fam Med, 37*(5), 360-363.
23. Yüksel, A. (2017). A critique of "Response Bias" in the tourism, travel and hospitality research. *Tourism Management*, *59*, 376-384.