Proceedings of the Twentieth International AAAI Conference on Web and Social Media (ICWSM 2026)

Pay to Win: Copyright Infringement and Academic Integrity Violation in Crowdsourced Document-Sharing Ecosystems

Margie Ruffin[1], Tzu-Bin Yan[2], Mutmaina Adebayo[3], Gang Wang[3], Kirill Levchenko[3]
[1]Spelman College, USA
[2]National Tsing Hua University, Taiwan
[3]University of Illinois at Urbana-Champaign, USA
margieruffin@spelman.edu, tbyan@cs.nthu.edu.tw, {adebayo5, gangw, klevchen}@illinois.edu

Abstract

Recently, crowdsourced academic document-sharing websites like Course Hero have been arousing controversy among media, students, and faculty. In this paper, we investigate the growth of these platforms and how they infringe on the copyright of university educators and aid in academic integrity violations of students. We collected and analyzed the metadata of over 227,000 documents shared on three popular platforms, including Course Hero, Docsity, and StuDocu, and found that Course Hero was the largest. Focusing on Course Hero, we then conducted two user studies ($n = 56$) and ($n = 153$) with university educators from the United States (U.S.). The first study analyzed Course Hero's impact on copyright and academic integrity violations. The second study examined the motivation behind faculty affiliation with Course Hero as "Verified Educators." Our results show that Course Hero has been growing exponentially since its inception. The results also highlight educators' general opposition to students' use of these platforms to gain an unfair advantage in coursework while acknowledging some alternative legitimate uses of these platforms. We show that educators create their own profiles on these platforms to identify copyrighted materials or initiate document takedown requests. However, educators do not always have copyright claims on concerning documents, and the takedown process is considered overwhelming. We conclude by discussing the implications of our findings and directions for future research.

1 Introduction

Crowdsourced academic document-sharing sites[1], such as Course Hero, host academic course materials and sell access to them on a subscription basis. In addition to a paid subscription, their users can also upload documents to earn credit for future downloads. (As of this writing, the front page of coursehero.com exhorts visitors "Upload your document to start studying.") Unfortunately, sometimes students upload course materials that they did not author themselves, in violation of the true author's copyright (Peiser 2022; Kolowich 2009). In many cases, what students upload are course materials produced by instructors—everything from syllabi, slides, assignments, and exams.

[1]In the paper, we will use "document-sharing sites" for short.

Some of the uploaded documents may be available publicly—via the course website—while others, such as exams, may not. As it is common practice for instructors to re-use exam questions, the availability of past exams poses a risk to *academic integrity* (McKenzie 2018). Indeed, while document-sharing sites like Course Hero, Docsity, and StuDocu claim to provide study resources for students, they have faced backlash for facilitating students gaining an unfair advantage in university courses (Gillis 2019).

Despite their dubious reputation, trading in course materials is big business. Course Hero, one of the platforms examined in this study, reported annual revenue in excess of $100 million in 2020 (Mascarenhas 2020). As a booming business, more websites are joining this track. Another recently founded platform, Docsity, reached $15.1 million in annual revenue in 2024 (RocketReach 2025).

Crowdsourced academic document-sharing sites are shielded from legal liability by disclaiming responsibility for uploaded content, placing the burden on users to ensure compliance with copyright law, as outlined in their Terms of Service (Hero 2025b). If copyright terms are violated, the burden falls on the copyright holder to file a Digital Millennium Copyright Act (DMCA) takedown request (Seng 2014), a process put in place in 1998 that allows Internet users to report copyright infringement. Unfortunately, a DMCA takedown can be a time-consuming and inconvenient process (Gillis 2019).

Motivation and Research Questions. Crowdsourced academic document-sharing sites have a distinctly bad reputation among many academics. In this study, we set out to rigorously investigate several aspects of this reputation. We started with the simple question,

- **RQ1**: What kinds of documents are hosted on crowdsourced academic document-sharing sites, and how fast do these platforms grow?

To keep the study tractable, we focused on materials from two closely related disciplines, Computer Science (CS) and Electrical and Computer Engineering (ECE), and 22 universities in the United States (U.S.). Among the three sites we examined—Course Hero, Docsity, and StuDocu—we found that Course Hero had the largest set of documents matching our criteria. Therefore, we restricted the rest of the study to Course Hero only. Next, aiming at the academics' sore spot,

we set out to determine:

- **RQ2**: How much of the materials (restricted to two disciplines and 22 universities) found on Course Hero are hosted without the copyright owner's permission?
- **RQ3**: How much of the materials (restricted to two disciplines and 22 universities) found on Course Hero do educators believe offer unfair advantages to students?

During the course of our investigation, we found that Course Hero appeared to have garnered faculty endorsements. In addition to the regular student accounts, Course Hero allows faculty members to create "Verified Educator Profiles" to "find teaching resources, discover innovative practices, and get recognition for their work." Importantly, Course Hero has been using the number of verified educator profiles on their platform to advertise faculty endorsement (Hero 2025d; Lederman 2020). In this study, we surveyed verified university faculty listed on Course Hero's site to determine,

- **RQ4**: How many university faculty choose to affiliate with Course Hero, and why?

Our Approaches. We started with RQ1 by collecting metadata from three academic document-sharing platforms—Course Hero, Docsity, and StuDocu—collecting a total of 227,606 documents across 22 universities. Among these, Course Hero emerged as the largest platform, prompting us to focus on its data when designing our follow-up user studies. We also crawled additional documents on Course Hero to estimate its growth trend over time.

After analyzing the metadata, we designed two surveys ($N = 56$ and $N = 153$) targeting university educators. First, for RQ2 and RQ3, we designed a survey focusing on specific faculty and their courses, asking them to reflect on their course policies, DMCA usage, and the potential impacts of Course Hero's documents on copyright and academic integrity violations. Note that this survey was *individually customized* for each surveyed educator by only showing specific Course Hero documents found about their courses. Then, for RQ4, we designed a second survey, asking faculty about their use of Course Hero and their motivation for the creation of "verified educator profiles."

Key Findings. Our research led to several important findings. *First,* we found that Course Hero, since its inception in 2007, had been growing *exponentially* and now hosts over 217 million documents. *Second,* through educators' assessment of their own course materials on Course Hero, we confirmed that Course Hero was indeed hosting unauthorized documents with instructor copyright, and a portion of such documents would give students an unfair advantage in their classes. *Third,* educators expressed their opposition to their students' use of Course Hero and particularly expressed concerns about documents *with solutions*. However, educators did not always have a copyright claim on such documents (e.g., some are students' work) to initiate DMCA takedown requests. *Fourth,* we found many educators were affiliated with Course Hero, but not in the way of positive endorsement. The majority of the educators with profiles made them to catch their students cheating or to perform DMCA takedown requests.

This research provides new insights into educators' perceptions of the impact of academic document-sharing platforms on copyright infringement and academic integrity violations. Our findings suggest that these platforms can be harmful to the original authors of the uploaded work and undermine academic policies designed to protect students' education. Further research is needed to explore the long-term effects of such platforms.

Contributions. We have three main contributions.

- Through empirical measurements, we revealed the *exponential* growth trend of Course Hero for crowdsourced academic document sharing.
- Our first user study revealed faculty perception of these websites and their impact on copyright and academic integrity violations.
- Our second user study provided new insights into why faculty choose to be affiliated with Course Hero as "verified educators."

2 Background and Related Work

Document Sharing in Higher Education. While students have been known to use different methods to gain an unfair advantage in their school courses (Clarke and Lancaster 2006; Eaton et al. 2019), crowdsourced academic document-sharing websites have emerged as a new contributor to the controversy. Although websites such as Chegg, Docsity, Quizlet, Course Hero, and StuDocu can be beneficial to students in need of extra resources, these platforms are being scrutinized by educators and administrators for their role in an uptick in academic integrity and copyright violations (Lowry 2020; McKenzie 2018).

The Pay-to-Win Business Model. Crowdsourced academic document-sharing websites such as Course Hero employ a "buy, sell, or trade" system to encourage students to share and download high-value materials (Slade 2021). They give students the option to either "buy" a monthly, quarterly, or yearly subscription or upload ("sell/trade") their own documents to the websites in exchange for the ability to "unlock" already uploaded materials (Dixon, George, and Carr 2021). This business model encourages students to upload course content, theirs or otherwise, in order to view documents they may need. Premier subscriptions on Course Hero can range anywhere from $39.95 per month per user to $119.40 per user for a one-year membership charged at once. For that price, students are only able to unlock up to 30 documents (Hero 2025c). Once they run out of "unlocks", students must turn to uploading more documents for continued access. There is minimal incentive and insufficient content moderation to encourage students to upload their own study materials, so they resort to paying for a subscription, which not all may be able to afford, or uploading materials that do not belong to them, causing undue harm to educators.

Copyright and Fair Use. File-sharing is not a new mechanism (Shami, Muller, and Millen 2011; Farnham, Turski, and Halai 2021), as it has been around for decades since the rise of popular peer-to-peer (P2P) file-sharing platforms

such as "The Pirate Bay" (Carrier 2009). Using P2P protocols like BitTorrent or Cyberlockers, Internet users may share files with little regard for legal repercussions (Kash et al. 2012). However, the downfall of these platforms came quickly when disrupted by outside forces, such as complaints and/or lawsuits from original copyright holders (Van Wynsberghe and Van Der Ham 2015; Ibosiola et al. 2019). As a result, P2P networks tried to integrate copyright protections into their protocols (Lou and Hwang 2009; Rodriguez, Tan, and Gkantsidis 2006; Kalker et al. 2004).

Crowdsourced academic document-sharing platforms thrive in a similar gray area as the older file-sharing platforms. It is possible students may not know what academic materials are copyright-protected (Gillis 2019; Rodriguez, Greer, and Shipman 2014; Fiesler, Feuston, and Bruckman 2015). The intellectual property (IP) ownership of academic content can vary by institutions (Hayashi and Fisher-Adams 2012; Holmes and Levin 2000) and without these policies being made clear to students, they may access these materials not thinking about possible risks but only about the perceived benefits of file sharing (Zhang 2015).

In this work, under U.S. copyright law, instructors may claim copyright if (1) they are the original author, or (2) they have made substantial changes to the work (Office 2020, 2025). With written policies stating that copyrighted content uploaded to their platforms without prior authorization is restricted, services like Course Hero and StuDocu can be released from liability for copyright violations so long as they provide a way for an educator to file a Digital Millennium Copyright Act (DMCA) takedown request (Hayashi and Fisher-Adams 2012; Hero 2025a; Studocu 2023; Ltd. 2024b; Seng 2014; Fiesler 2020). The DMCA takedown system was put in place in 1998 to provide legal protections against unauthorized access to copyrighted works on the Internet (Ltd. 2024a).

Combating Academic Integrity Violations Online. Certain academic content, such as syllabi and lecture slides (Dixon, George, and Carr 2021), may seem harmless in aiding cheating, while materials such as exams, quizzes, code, and lab reports can be more problematic (Lieneck and Esparza 2018). Maintaining academic integrity in the age of information sharing is challenging (Dyer 2010), as researchers have tried both manual and automated methods to monitor document-sharing platforms like Course Hero (Dixon and George 2021). Prior work by Dixon et al. is focused on a small sample, generating 55 reports from 386 unique documents using Google alerts based on keywords (Dixon and George 2021). Some work goes as far as attempting to automatically monitor document uploads to more than just Course Hero (i.e., Stack Exchange and Chegg) (Somers et al. 2024). Our work aims to provide a more comprehensive understanding of the volume and growth rate of documents shared on these platforms. More importantly, we seek to understand how educators view and address the issues of copyright infringement and academic integrity violations.

School	Course Hero	Docsity	StuDocu
University of Texas at Austin	369,684	6,337	11,330
Arizona State University	326,945	2,625	19,899
University of California, Berkeley	315,615	5,027	10,062
University of Illinois U-C	255,528	8,476	12,518
University of Florida	225,200	7,328	8,878
University of California, Davis	202,855	4,715	7,178
Georgia Institute of Technology	174,025	9,391	12,917
University of Arizona	144,209	4,062	9,340
University of Washington	133,243	7,698	6,647
Cornell University	118,304	1,028	6,296
Stanford University	113,666	2,885	4,850
Boston University	99,614	105	7,725
North Carolina State University	92,636	3,260	5,142
Stony Brook University	89,281	25	9,891
MIT	57,457	2,577	4,958
University of Utah	45,625	2,545	5,023
Carnegie Mellon University	36,361	508	2,976
Wash. University in St. Louis	26,526	17	3,874
UC Santa Cruz	25,188	1,539	2,153
Princeton University	9,369	900	1,588
California Institute of Technology	5,476	308	274
Dartmouth College	5,317	149	518
Total	**2,872,124**	**71,505**	**154,037**

Table 1: Total document count on Course Hero, Docsity, and StuDocu, across *all disciplines*, for the 22 selected schools.

3 Methodology

To answer our research questions, we analyzed the crowdsourced academic document-sharing ecosystem, focusing on a sampled set of schools and their Computer Science (CS) and Electrical Computer Engineering (ECE) departments. We targeted CS and ECE due to our familiarity with the subject matter (we did not evaluate other subjects as we would be unable to assess their content accurately). We chose these departments also because of the significant number of related documents on these platforms. We selected 22 institutions, given our limited budget/manpower for data collection and user study construction (detailed justifications later). We first consider the 10 U.S. universities with top-ranked CS programs by the U.S. News & World Report (2022) (L.P. 2022). However, we did not want to focus solely on the highest ranking programs, so to diversify, we also consider those ranked 40th–50th at the time[2]. Table 1 lists the selected schools. We ended up with more than 20 schools because some schools share the same rank.

After collecting the document datasets, we conducted two user studies with *university educators* to (1) identify unauthorized, copyrighted documents that they think could give students an unfair advantage and (2) understand their motivations for using these platforms. We chose *not* to survey students to avoid potential harms and risks (e.g., re-identifying student users of Course Hero, linking students to academic integrity violations, and getting dishonest survey responses). The first *Rights and Integrity Survey* cov-

[2]We did not include the University of Notre Dame (ranked within the 40th–50th range) because we observed that a noticeable amount of documents belonging to the Notre Dame school on Course Hero actually belong to a few international schools, not the U.S. University of Notre Dame in Notre Dame, IN.

ered course procedures and specific documents from Course Hero. In the second *Verified Educators Survey*, participants shared insights into their affiliation with Course Hero.

Ideally, including more schools would further increase the data diversity. We limit our study to this number of schools and departments for two main reasons. First, fully automating the data collection from these document-sharing platforms is challenging. A great deal of work had to be done manually, and we did not have the resources to scale up the effort. Our data crawling budget is restricted by the rate limit of the browser animation (of the crawler), the number of IPs and machines we can control, and the resources needed for CAPTCHA solving. The datasets of the 22 schools already took the team 6 months to collect. Second, our survey (more specifically, the Rights and Integrity Survey) needed to be individually customized for each of the surveyed educators since we needed to display the specific documents collected about their course. It took an additional 2 months to manually customize the Course Hero document statistics pages, and the survey pages for *each class* and their educators. Examples of these customizations can be seen in our supplementary materials (Ruffin et al. 2026). For these reasons, we did not sample more schools but tried to diversify the schools within our limited resources for data collection.

3.1 Measurement Method

Document Collection. Following an in-depth Internet search, we determined that the three most popular academic document-sharing websites in the U.S. were Course Hero (Hero 2025b), Docsity (Docsity 2024), and StuDocu (Studocu 2025). Other websites we found did not follow the same document upload structure (Quizlet 2025; Chegg 2025) or did not organize documents based on schools (DocSlides 2025; StudyLib 2025). We did not consider sites such as Chegg, for example, because they focus on answering student questions (e.g., homework questions from textbooks) rather than collecting academic documents. Table 1 shows the total number of documents for each school across *all disciplines* (not just CS and ECE) on each of the three platforms (these numbers are self-reported on these websites). Although CS and ECE documents comprise a smaller subset of the total, as shown later in the paper, they greatly impact copyright protections and academic integrity violations.

To better understand the selected platforms, we developed several Selenium-based web crawlers (Muthukadan 2025) to gather document metadata from our target departments and schools. Each document (i.e., a file) has metadata such as a title, a short description, the uploader's information, upload time, number of views, document type, and its course number. Ethical considerations are addressed in Appendix A. In total, we gathered 227,606 documents from the three platforms related to the CS and ECE courses. As reported in Table 1, the total number of documents across *all disciplines* from these target schools is 2,872,124. The CS and ECE documents take about 6.5%. Course Hero emerged as the most popular platform (with 225,785 documents collected), surpassing StuDocu and Docsity. It was clear that Course Hero, the largest website of this kind in the U.S,. had become a *monopoly* that dominated the market. Therefore, we based our user studies on Course Hero data.

Growth Estimation. After analyzing the document metadata from Course Hero, we found evidence to support the idea that the *document_id* tag correlates with the upload time on Course Hero. These IDs appear sequential, starting from as low as 8 in 2007 (when Course Hero was founded) to over 217 million in 2024, incrementing by 1. To measure the platform's growth, we conducted a week-long intensive data crawling, pinging document URLs incrementally to collect 10,000 data points. Of these, 2,308 (23.08%) were missing, likely due to duplicate, irrelevant content, or someone having submitted a DMCA takedown request. A second random crawl of about 5,000 data points, targeting *document_id* values between 10,000 and 217 million (the largest ID we could find), revealed 1,476 missing documents (26.83%), further suggesting internal or external content moderation. This extra dataset (of *document_id*s and timestamps) allows us to estimate the growth trend of Course Hero, and the result is presented in Section 4.1.

3.2 Survey Design

The two surveys seek to understand (1) educators' perception of copyright and academic integrity violations present on Course Hero, and (2) the motivation behind their affiliation with "verified educator profiles" on Course Hero. In our study, an educator is an individual currently teaching (or has taught) a college/university course. In the following, we describe the study process from the participants' perspective and explain our design choices. Both studies were conducted under our Institutional Review Board (IRB) approval, with minimal risk to the participants (ethics discussion in Appendix A. The question lists are in Appendix B and C.

Rights and Integrity Survey. The first survey contains two parts and an exit question, as shown in Figure 1. We target university instructors because we believe only the instructors can help identify the copyright owner of the documents related to their courses, determine whether the documents are uploaded with permission, and articulate whether the documents can give students an unfair advantage in the class. The detailed questions are listed in Appendix B. If not otherwise stated, the question options are "Yes", "No", "I am not sure" and "Prefer not to say." Below, we describe the survey workflow using Figure 1.

General Questions. Participants are first instructed to read and sign the consent form and are then given a brief introduction. Following this, they start with step ❶ of Figure 1, where they are asked several questions about their course practices and their knowledge and use of the Digital Millennium Copyright Act (DMCA), see Table 3 in the Appendix. These questions are used to understand if faculty have strict or relaxed policies around document-sharing, their knowledge of the document-sharing platforms, and how much effort they are willing to expend to take down unauthorized uploaded documents.

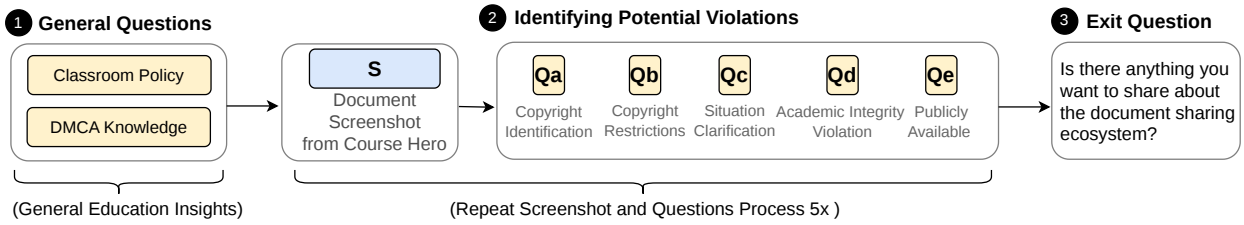

Figure 1: We show the study process from *a participant's perspective* — In step ❶, the participant answers several general to understand their classroom policies and knowledge of DMCA. In step ❷, the participant examines 5 documents on Course Hero related to their course and answers questions related to them. In step ❸, the participant answers one additional exit question.

Copyright and Academic Integrity Violation. This portion of the survey is *individually customized* for each surveyed instructor. In other words, each surveyed instructor would only review the documents found on Course Hero related to their own courses. This level of customization is *necessary* because we cannot expect instructors to reliably assess documents from other people's courses. Only by showing them their own course documents can they help to determine the copyright owners of the documents and judge whether the documents give an unfair advantage to students in their class. To this end, we hand-craft the survey page for each individual surveyed instructor (which is unfortunately not scalable).

In step ❷, participants review five *screenshots* of documents from Course Hero for their course. Each screenshot only shows the first page of a document. We select documents with the highest average daily views per document type. For courses with more than five document types, we choose the most-viewed documents. For courses with fewer than five document types, we repeat the selection until reaching five. This method ensures both diversity and relevance. Personally identifiable information (PII) in the screenshots (e.g., uploader names) was redacted to preserve privacy.

Under each document, participants answer five questions (see Appendix B for the full questions and answer options). The first question **[Qa1]** asks about the document's authorship to identify copyright holders. If they indicate they are not the original author but have contributed to its creation, a follow-up question **[Qa2]** asks if any usage restrictions are placed by the original author. After that, we ask the participants how the document was uploaded to the website and what they would like to see happen to it in **[Qa3]**. This helps to inform whether the instructor uploaded the document themselves, whether they approved it to be uploaded, and whether they may want it taken down.

Finally, we ask about the possibility of the document giving students an unfair advantage **[Qa4]** and its accessibility to the public **[Qa5]** (i.e., whether students can access it without using Course Hero).

Exit Question. In the last step ❸, we asked: *Is there anything that you want to share with the research team about your feelings concerning the questions and the document-sharing ecosystem in general?* This gives the faculty an opportunity to share additional thoughts they may have about our study and the document-sharing ecosystem in general.

Verified Educators Survey. Participants for this survey are university educators who have a *Verified Educator Profile* on Course Hero. After participants read and sign the consent form and read a brief introduction to the study, they are instructed to answer a single question, which will potentially be followed by more, depending on how they answer (see Appendix C for full questions and answer options). The first question **[Qb1]** aims to understand how many of Course Hero's verified educator profiles are legitimately created. If they answer "yes", Qb2 and Qb3 will be displayed. The next question **[Qb2]** helps gauge the level of interaction with the platform through profile creation, and **[Qb3]** helps to understand the ways in which a participant may want to engage with the platform. Finally, **[Qb4]** is used to validate the use of the profile if they indicated their reason for making it was to remove documents from the website. For most questions in this survey, participants select from a list of single-choice or multi-choice answers unless otherwise stated.

3.3 Recruitment

We recruit participants for both studies based on the data gathered from Course Hero. Because the surveys differ in context, we use different recruitment methods to reach our target audiences. For both studies, educators reviewed and acknowledged consent before taking the survey. They could request data removal at any time, and all PII was anonymized before data analysis. Data was stored securely on a campus cloud server.

Rights and Integrity Survey. This survey is customized for educators from the 23 selected schools in Table 1. For each school, we prioritized popular courses (with a large number of documents on Course Hero) and selected instructors who had taught the course recently or frequently for greater insights into the documents found. We conducted a manual Internet search using the school and course information to find the professors and their emails.

Out of 123 emails sent, we received 57 responses (a 42.42% response rate). We were only able to send 132 emails because the survey needed to be personalized, including customized survey pages for each course and individualized survey links for each instructor we contacted. The customization was done manually. Example customized survey pages and questions can be found in our supplementary materials (Ruffin et al. 2026). Two responses were incom-

plete: one was removed due to only partially answered general questions. The other one contained complete copyright answers and thus was included in the analysis. In total, we had $N = 55$ valid responses for part one and $N = 56$ for part two. The Rights and Integrity survey took a median of eight minutes to complete.

Verified Educators Survey. This survey targeted educators who have a "verified educator profile" on Course Hero. We started by randomly sampling survey candidates from a list of verified educator profiles located in the U.S. To deliver the survey request, we would need their email addresses, which were not directly available on the educator profiles. As such, we conducted a manual Internet search for educator emails using their names and the accompanying information on their profiles (such as their title, school, and department) as additional search keywords.

Out of the 2,622 sampled educators, we found emails for 1,013 to send the survey request. From the 1,013 emails sent, 676 universities/colleges/trade schools across the country were represented. Nine of these were invalid or belonged to individuals no longer in the listed position on Course Hero, and thus, they were excluded. From the 1,013 emails sent, we received 155 responses (a 15.33% response rate). Two responses were incomplete and filtered out, leaving us with a total of $N = 153$ valid responses for the study. The Verified Educators Survey took a median of two minutes to complete.

4 Results

4.1 Course Hero Document Analysis

We start with **RQ1** to analyze the documents in Course Hero and the growth trend of the platform.

Size and Growth. Using the method described in Section 3.1, we estimated the size and growth of Course Hero based on its sequential document IDs. As shown in Figure 2, document growth on Course Hero from 2007 to 2024 has been *exponential*. Based on the largest *document_id*, we estimate that the platform has hosted more than 200 million documents. However, not all documents are accessible. We crawled 15,500 documents to estimate the document space density, finding 11,716 available (75.59%). This number fluctuates due to content removals, such as DMCA takedowns or irrelevance. A close look at Figure 2 also reveals an excess of data points under the curve or "backfilling", where newer documents are assigned lower IDs, likely to fill gaps left by removed documents.

Document Breakdown. To analyze the types of documents on Course Hero, we focused on Computer Science (CS) and Electrical Computer Engineering (ECE) as these are our areas of expertise. For the rest of the analysis in this paper, we focus on documents from these two subject areas. We crawled metadata for CS and ECE from 22 selected schools (Table 1). For Dartmouth, we only obtained CS data since they don't have an ECE Department comparable to other universities.

We collected metadata for 225,785 unique Course Hero documents (CS: 126,833, ECE: 98,952), which are categorized upon upload. Here, we want to provide a brief context for the dataset size. Based on our estimation above, Course Hero has more than 200 million documents in total. If we only consider the 22 selected schools (all disciplines), as seen in Table 1, the number of documents is 2.87 million. If we only consider CS/ECE disciplines in the selected schools, the dataset contains about 225,000 documents. Between January and April 2023, Course Hero reorganized its document classification into categories like *Assignment, Code, Essay, Lab, Notes, and Slides.* While this reorganization improved classification accuracy, we found the labeling still imprecise. To address this, we manually labeled a subset of documents to use in our first user study (Section 4.2). We share the corresponding results later in the paper.

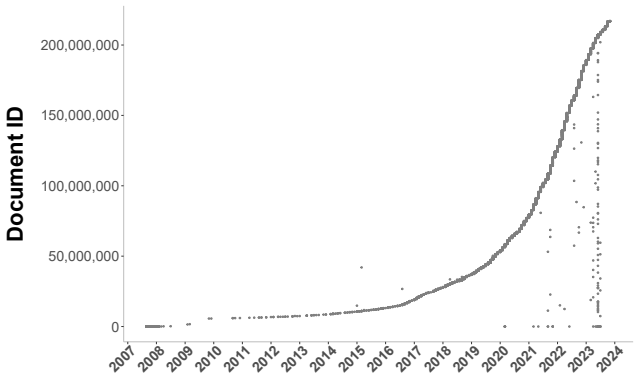

Figure 2: Document IDs spanning from Course Hero's inception in 2007 until 2024.

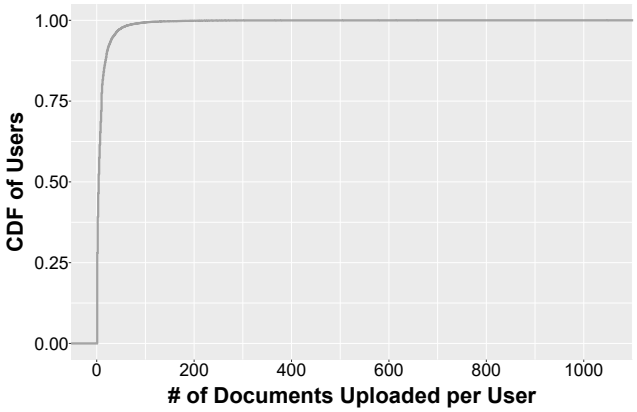

Figure 3: Cumulative distribution function (CDF) of the number of documents uploaded per user.

The metadata included tags such as *total_views* and *has_answers*. Within our data, the number of views for the documents combined is 7,296,737. While the total view count is high, only 12,271 documents (5.43%) are marked as "popular" by Course Hero's own standards, boasting more than 100 views individually. In total, 80,331 documents (35.58%) are reported to have answers.

The metadata also revealed 23,957 unique uploaders. As shown in Figure 3, the majority of uploaders (99.83%) captured in our metadata have uploaded fewer than 50 doc-

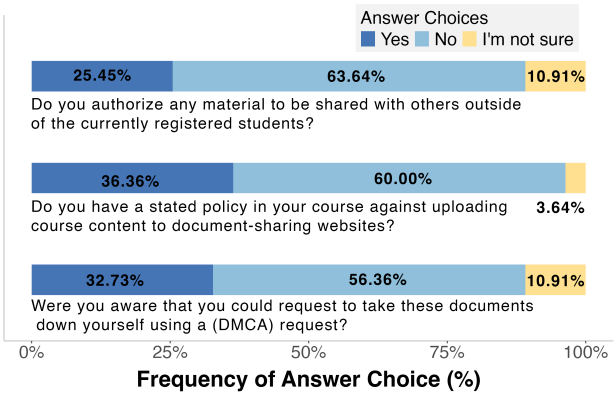

Figure 4: Educator responses to general questions from the Rights and Integrity Survey.

uments. Only a small portion of uploaders (140 users, 0.067%) have uploaded more than 100 documents. . The unique uploaders included both students and verified educators. Course Hero offers verified educators free access to course materials like assessments and lecture notes created and shared by faculty and students. Although their site claims over 200,000 educator profiles (Hero 2025d), many appear blank, containing only a name and an affiliation. We found that out of the 200K+ educator profiles on Course Hero, only 10,902 of them (5%) had one or more documents uploaded. Of these, 3,913 had only one document uploaded. This indicates that the vast majority of educator accounts did not upload documents to the site. This observation led to our question (**RQ4**): why are university faculty affiliated with (have profiles on) Course Hero?

SUMMARY: (1) Since its inception in 2007, Course Hero has grown *exponentially*, now hosting over 200 Million documents across schools and disciplines. (2) Document popularity is skewed: a small percentage of documents (5.43%) get 100+ views.

4.2 Rights and Integrity Survey

Next, we explore **RQ2** and **RQ3** related to copyright infringement and academic integrity violations.

General Questions. The general questions are focused on course practices and DMCA awareness. Table 3 (in the Appendix) lists the full questions and their answer distribution. Figure 4 highlights a few important questions.

We start with class policies. With $N = 55$ participants, Figure 4 shows that 63.64% of educators oppose sharing course materials with students not enrolled in their classes. However, most educators (60.00%) did not include policies in their syllabi or course websites to explicitly prohibit this activity, as shown in Figure 4. By correlating the answers of these two questions, we find that 40% of the surveyed educators are in the intersection. Table 3 reveals that 98.18% of the surveyed educators have never uploaded their course content to platforms like Course Hero (consistent with our data analysis above in Section 4.1.) Overall, we show that educators oppose content sharing outside of the class but often do not explicitly inform students that it is prohibited. We could not infer the reason based on our survey results.

Table 3 shows educators are generally aware that their work is being shared online: 38.18% of them "have heard about course documents shared on doc sharing websites", 20.00% of them "have checked out the documents themselves", and 12.73% of them "have filled out a DMCA request to have them removed." However, there are still 21.82% of the educators *unaware* that their course content was shared on Course Hero. It is arguable that fewer educators have filled out DMCA requests because they simply did not know they could. As shown in Figure 4, when asked if they were aware that they could request for their course content to be removed from Course Hero, 56.36% of participants said "no". In other cases, faculty are less likely to fill out DMCA requests for more reasons, including "there's too many documents" (42.86%), "takes too much time" (42.86%), or "they do not care they are there" (42.86%).

Given the vast number of documents on Course Hero, filing a DMCA request for each suspected copyright violation can be overwhelming. We asked the participants how much effort they would be willing to expend to remove the documents (allowing them to select multiple options). While 12.73% were unwilling to take any action, the majority, 61.82%, were open to filling out a short form for all documents, and 43.64% were willing to perform a one-click action per document. This suggests most educators are open to taking a low-effort measure to take down course documents from Course Hero.

SUMMARY: (1) Educators expressed their opposition to the students' use of Course Hero, but they often failed to proactively inform students against uploading copyrighted materials. (2) Even with their opposition, the burden of taking the documents down themselves could seem so overwhelming that few educators chose to do so.

Labeling Document Categories. As described in Section 3, after answering the general questions, participants were asked to assess documents collected from Course Hero related to their courses. Participants answered questions related to the copyright owners of documents and whether the documents would help students gain an unfair advantage in their class. To provide the context for our analysis, we first need to label the document type/categories, considering Course Hero's document labels are unreliable. In total, 440 documents were used to set up the Rights and Integrity Survey. However, not all invited educators ended up taking the survey. The number of documents that were assessed by educators was 280. Two researchers manually coded a sample of 100 out of 440 documents (5 per course for 22 schools). After agreeing on 15 labels, the remaining 340 documents were divided and individually coded. Figure 9 (in the Appendix) shows the categories of the 440 documents labeled by us. Figure 5 shows the categories of the 280 documents assessed by educators. Assignments, Exam Materials, and Lab documents are the most common types of documents in

the collection. Figure 6 shows the average number of views per document under each category. Textbook, Lab, and Quiz documents are among the top that gain higher views.

Document Copyrights. We first analyze the participants' answers regarding document copyrights. According to the U.S. Copyright law, an instructor can make a copyright claim on a document if they are the original author or have made substantive changes to it, making it a derivative work (Office 2020).

Out of the 280 documents, the surveyed educators (instructors) had copyright claims on 37.15% of them. Of these, 21.45% documents had the surveyed educator as the sole author. 8.36% of the documents had the surveyed educator as one of the original authors, with others contributing. 7.64% had the educator as a contributor but not the original authors. Only 3.64% of the latter group faced restrictions from the original author. For the remaining documents (62.85%), the surveyed educators either did not have a copyright claim or they were unsure. For example, the uploaded documents (e.g., lab reports) could be entirely the work of students, and thus the copyright belonged to the students, not the instructor. In other cases, the copyright of the documents may belong to instructors, but *not the instructor that responded to our survey* (e.g., a former instructor who taught the class in the past). As such, the percentage of documents with instructor-copyright (37.15%) can only be interpreted as a *lower-bound*.

SUMMARY: Educators identified documents from Course Hero on which they had copyright claims (37.15%), confirming the copyright infringement of the platform.

Academic Integrity Violations. Under each document, we further asked the educators to assess whether the document would give students unfair advantages in their classes (e.g., helping students to cheat). Here, we examine the overlap between documents with instructor copyright claims and documents that give students an unfair advantage. The result is presented in Figure 7.

Out of 280 documents, 28.22% were marked as providing an unfair advantage to students (the right bar in Figure 7). Under this category, 12.86% were authored or co-authored by the educators, and 15.36% had undetermined copyrights. This "intersection" is important because only when the instructor has a copyright claim on the document can they initiate a DMCA takedown request. However, our result indicates that educators *do not always have a copyright claim* on the concerning documents that give students an unfair advantage (15.36%) to initiate DMCA takedown requests. These documents may be created by other educators or are those of the students' work.

We further analyze which types of documents are more likely to give students an unfair advantage. In Table 2 (Appendix), we show a number of documents (and their average views) under each document category and compare those that give students an unfair advantage with those that do not. We have two quick observations based on the descriptive statistics. First, Assignment, Code, and Exam Materials have the highest number of documents that give students an unfair advantage. This is aligned with intuition since these documents are often closely related to student assessment and student grades. Second, we observe that for most categories (except for Code and Professor Notes), documents that give students an advantage take a smaller portion compared with documents that do not give students an advantage. However, the smaller portion of "advantage" documents usually have a higher view count per document.

SUMMARY: (1) Educators confirmed 28.22% of the documents could help students gain an unfair advantage in their courses. Assignment, Code, and Exam Materials contained the highest number of such concerning documents. (2) Unfortunately, educators did not always have a copyright claim on such documents to initiate a DMCA takedown.

Other Concerns from Surveyed Educators. Before completing the survey, we asked participants if they had any additional thoughts to share. We conducted a thematic analysis (Braun and Clarke 2012) of all 39 valid responses using a codebook developed by two authors who independently coded the first 32 responses (Strauss and Corbin 1997). The initial inter-coder agreement, measured by Cohen's Kappa, was 0.93, indicating strong consistency. The two coders and remaining authors resolved any disagreements to finalize the codebook, which was then used to label the remaining responses. A single response could be assigned multiple codes. Table 4 (Appendix) shows the codebook.

Documents with "Solutions". Ten educators mentioned that they were particularly concerned about documents that included solutions. For example, **P7** explained, *"I have see[n] documents on course hero that do include solutions. Those are the ones I'm worried about."*. **P10** also detailed, *"I have seen where a current assignment is posted along with a solution. That is definitely the one type of document that we don't want, and will typically result in us making a request to take it down. Often, we don't find out until it is too late though."* Documents from years ago and from past instructors can be a concern if they include the solutions (**P23**): *"Those would provide a clear advantage to students who have access to them since some of the content is reused."*

Copyright and Integrity Violation. Many of the surveyed educators (26) further confirmed their negative sentiment towards academic document-sharing websites due to copyright infringement and academic integrity violations. For example, **P13** mentioned, *"Some of the documents are lecture notes, which I released under a CC-BY-NC license; CourseHero is violating that license by requiring payment to download those documents."* **P24** also shared that, *"Sharing these documents is a clear violation of the course policies, and provides an avenue for future students to cheat/plagiarize assignments."*

Alternative Document Sharing Platforms. The surveyed educators mentioned alternative channels that students used to share academic content. For instance, **P43** explained, *"At our university, because some students (frats, for example) have "back files" on courses, I have always published all*

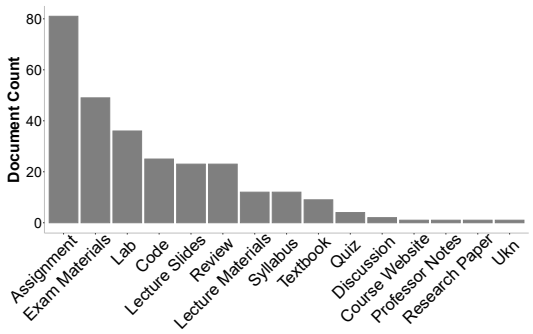

Figure 5: Categories of documents used in the Rights and Integrity Survey (280).

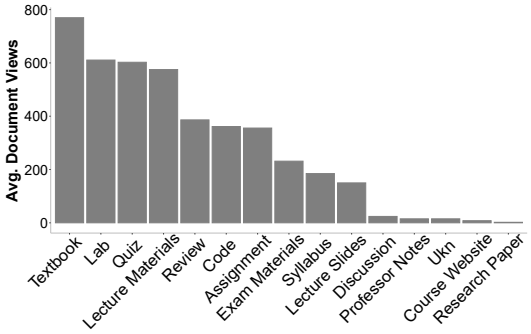

Figure 6: Average views per document for documents used in the Rights and Integrity Survey (280).

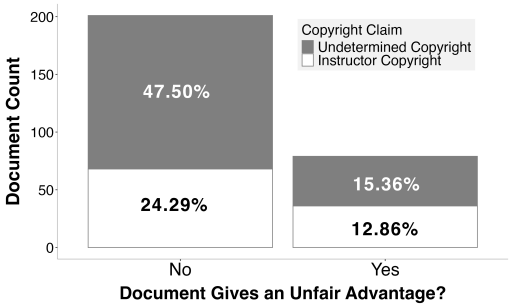

Figure 7: Comparison of documents with educator copyright vs those that give students an unfair advantage.

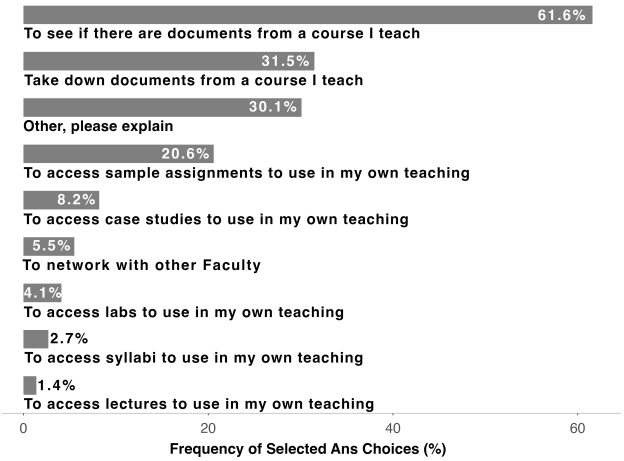

Figure 8: Educator responses on why they created Verified Educator Profiles on Course Hero.

materials from previous semesters for all to see." We also note that three different responses (**P21, P26, P29**) mentioned the sharing of assignments on GitHub as an issue in their courses.

SUMMARY: (1) Educators expressed their concerns about documents *with solutions*. (2) Course Hero is not the only platform that educators have to be worried about, as students find alternative ways to share course content.

4.3 Verified Educators Survey

Recalling Profile Creation. To investigate **RQ4**, we analyze user responses to the question regarding the creation of a verified educator profile on Course Hero in their name. Participants can choose one of the four options: "yes", "no", "I'm not sure", and "prefer not to say." We conservatively code "prefer not to say (NTS)" to the negative category (i.e., no). With $N = 153$ participants, 48.37% do recall creating their verified educator profiles on Course Hero. The remaining 51.63% of the participants either claimed "no" or could not remember creating a profile.

When creating a verified educator profile, users may choose to upload a profile photo. There were 11 educators out of 153 participants identified as having a photo on their profiles. Out of these 11 educators, only 7 indicated that they did create Course Hero profiles, and only 3 indicated they uploaded the photo themselves. 2 educators stated they did not upload the photo, and 2 stated they were not sure.

Reason for Profile Creation. As shown in Figure 8, the most common reason for educators to create their profiles was to "To see if there are documents from a course I teach" (61.6%). The second most popular answer choice (31.5%) was "To take down documents from a course I teach," followed by "To access sample assignments to use in my own teaching" (20.6%). Other less popular selection options include "To access case studies/lectures/labs/syllabi to use in my own teaching" and "To network with other Faculty", totaling 21.9%. We perform a thematic analysis (Braun and Clarke 2012) on all 22 (30.1%) responses from those who selected *Other, please explain*. To develop a code book, two authors independently coded the responses (Strauss and Corbin 1997). The inter-coder agreement via Cohen's Kappa is 0.95, indicating a high agreement level. Two coders and the remaining authors resolved the disagreement. Table 5 (Appendix) shows the code book. The codes can be divided into "check for illegitimate use" and "instructional use."

A: Check for Illegitimate Use. This is the most com-

mon theme. Five educators explained that they created their profiles to specifically check for academic integrity violations and prove Course Hero is a source of them. For example, **P14** responded *"I think it's a cheater's resource and I wanted to see what was there. I'm thinking about uploading bogus solutions to see if I can catch students abusing the site."* Moreover, **P126** explained, *"I had encountered students turning in work identical to each other. I searched for the exact phrases used and found search results to Course Hero. I created an account to be able to view what is available for the course I teach so I could prove the students were plagiarizing from Course Hero."* Three educators chose to create accounts to see if their work was on the platform and to perform DMCA takedowns. One faculty member felt they needed to create an account so they could *"remove test questions that had been posted by a student* **(P12)**." Another educator expressed the need to "*stop CourseHero from using my copyrighted materials without [their] permission* **(P42)**."

B: Instructional Use. Three faculty members specified that they created accounts to check for syllabi for course transfers. For example, participant **P47** explained how they use Course Hero to *"verify [the] articulation for the eligibility requirements for [their programs]. [They] use Course Hero to review the syllabi of courses in question to ensure they are covering the appropriate topics required by our programs (because prospective applicants have taken these courses)."* Others mentioned they created accounts to download content for homeschooling or just to attend Course Hero conferences.

SUMMARY: (1) The *vast majority* of the surveyed educators made their verified educator accounts to catch their students cheating or to perform DMCA takedown requests. (2) One alternative reason for making an account was to check syllabi for course transfers. The result suggests that "verified educator profiles" should not be treated as a sign of endorsement from faculty members (but could be the opposite).

5 Discussion

Findings and Implications. The main contribution of our paper is to understand the *educators' perception* of crowdsourced academic document-sharing websites (Course Hero in particular). By surveying educators and asking them to assess documents from their own courses found on Course Hero, we obtained concrete evidence that (1) Course Hero indeed hosts documents with instructor-owned copyrights, and (2) some of the documents can give students an unfair advantage in classes (e.g., helping students cheat). In addition, Course Hero has been promoting their affiliated educators and using the number of "verified educator profiles" on the platform to give the impression/illusion of faculty endorsement (Hero 2025d; Lederman 2020). However, our study reveals that the majority of the surveyed educators created the verified educator profile/account in order to catch their students cheating or to perform DMCA takedown requests. As such, the verified educator profiles should not be interpreted as a form of endorsement from educators (which may even represent the opposite of endorsement). Given the *exponential growth* of websites like Course Hero, we believe further actions are needed to address these concerns.

There Are Some Legitimate Uses But Are They Enough? Our results show that some of the uses of document-sharing platforms may not harm students or educators. For instance, faculty use posted syllabi for course transfers, and parents use content to support homeschooling. These positive uses could help reduce the negative perception of such platforms. Future research should explore how these platforms would function if they removed both unauthorized (copyright-protected) content and documents that aid academic integrity violations while focusing on these legitimate uses. These aforementioned implications raise the question: would educator-authorized content alone be enough to keep the platforms valuable and appealing?

Awareness, Perception, and Actions. Throughout our survey, we found various "gaps" between the awareness, perception, and actions of educators with respect to unauthorized content sharing. For example, some of the surveyed educators (21.82%) were *unaware* that their course content was on Course Hero. In addition, most educators (56.36%) said "No" when asked if they were aware that they could request to take documents down via DMCA. Even though most educators expressed their opposition to content sharing, such *perception* may not always lead to *actions*. For example, most of them (60%) did not take proactive measures to inform students or prevent them from doing so, e.g., in the form of course policies (actions). Similarly, only 20% of the surveyed educators have submitted DMCA requests themselves, primarily because there are many documents on Course Hero related to their course, and it is time-consuming to submit a DMCA request for each of the documents.

To address these issues, while it is important to raise the awareness of educators and develop tools to ease the process of copyright management, it may be insufficient without the cooperation of document-sharing platforms. Currently, Course Hero has made it clear in their Terms of Use (Hero 2025b) that uploaders should not "*[upload] any Content not owned by [them] (i) without the prior written consent of the owner of that content or (ii) in a way that violates someone else's (including Course Hero's) rights*". However, Course Hero has also stated that they are "*not responsible for ... screening or monitoring ... any content.*" Our result shows that the current practice from Course Hero is insufficient as most users probably do not read the long document of Terms of Use. A potential improvement from the Course Hero side is to present a clear warning message to the uploader *at the time of document upload* to discourage them from uploading unauthorized documents.

Academic Integrity vs. GenAI. In the past few years, the rise of Generative Artificial Intelligence (GenAI) has posed new challenges to academic integrity in higher education. In particular, Large Language Models (LLMs), a type of GenAI (e.g., OpenAI's GPT-4 (ChatGPT 2025) and Anthropic's Claude AI (Claude 2025)), have already substantially impacted education (Kumar et al. 2020). Students have started to use them to help with their homework, lab assign-

ments, and even their literary writing assignments (Meça and Shkëlzeni 2024). Schools nationwide are forced to recognize the benefits and potential harms of having GenAIs in the classroom. As GenAI can be used for both assisting student learning and facilitating cheating, it is increasingly difficult to define and detect academic integrity violations that involve the use of AI (Perkins 2023). While GenAI is not the focus of this paper, our study has a few implications. First, the large number of documents (e.g., exams, homeworks) accumulated on academic document-sharing websites like Course Hero is an attractive data source for *training* such GenAI models. However, given the way in which these documents are collected, training AI models on such data would raise further copyright questions. Second, with the rise of GenAI, it would be interesting to see whether the use of academic document-sharing websites would start to decline in the near future. From our current data (e.g., Section 4.1), we have not yet observed such a trend, as academic document-sharing websites are still growing.

Our Recommendations. Taking a *proactive* approach against unauthorized sharing of educator-authored content may help reduce its unwanted distribution. We recommend that educators clearly inform students, both verbally and in writing (in syllabi or on course websites), that sharing course content is not authorized. Additionally, course documents could include a specialized notice, similar to a copyright notice, prohibiting redistribution. This could simplify future disputes if a DMCA takedown request is needed. In our survey, we asked how much effort educators would be willing to expend to remove their documents and found an *automated DMCA request process* could ease the burden of managing multiple requests. Finally, it is important to raise the *awareness* of educators regarding unauthorized document-sharing activities online as well as their rights to take down documents via DMCA. This effort can be potentially integrated into the current Teaching Assistant (TA) training and Faculty Onboarding process.

Limitations and Future Work. We note several limitations to our work and discuss future directions of research. *First,* we only focus on Course Hero's college-level CS and ECE courses for 22 schools. Future work may consider including more schools (or even high schools) and different academic fields to study their differences and examine whether our observations generalize. *Second,* the schools selected are U.S.-based institutions. Course Hero and other platforms host documents from schools all over the world. Perceptions may differ in other regions and would need to be further studied. *Third,* our user study responses may have been limited due to a self-selection bias. Those educators who responded to our survey requests may have been more likely to care about related issues. *Fourth,* we did not survey students in this work; future work may explore this direction with careful designs to control/reduce potential risks/harms (to students). *Finally*, in this work, we surveyed educators, asking them to manually assess the copyright owner of a document and determine whether it can provide an unfair advantage to students. This process was time-consuming and highly unscalable. Future work may look into ways to *automate* this process (e.g., using Large Language Models), which can be helpful to narrow down a smaller set of documents of interest for manual inspection and/or request DMCA takedowns.

6 Conclusion

Through extensive data collection and two user studies, we analyze how crowdsourced academic document-sharing platforms impact educator copyrights and student academic integrity. We find that the largest platform, Course Hero, hosts over 200 million course documents with little content moderation and has been growing exponentially. Educators are able to identify copyright-protected documents on Course Hero that are shared without permission. They also confirm that certain documents (especially those with solutions) can give students an unfair advantage in their courses. Finally, we show that most faculty members surveyed created "Verified Educator" profiles to identify copyrighted materials or initiate document takedown requests. Our result raises new questions for future research regarding the effects of these platforms on higher education and student learning.

Acknowledgments

This work was supported by the NSF Graduate Research Fellowship Program under Grant No. 21-46756. It was also supported by the National Science and Technology Council (NSTC, Taiwan) grant NSTC 115-2222-E-007-004.

References

Aziz, M. A. B.; and Wilson, C. 2024. Johnny Still Can't Opt-out: Assessing the IAB CCPA Compliance Framework. In *Proc. of PETS*.

Braun, V.; and Clarke, V. 2012. *Thematic analysis*.

Carrier, M. A. 2009. The Pirate Bay, Grokster, and Google. *Journal of Intellectual Property Rights*.

ChatGPT. 2025. ChatGPT | Introducing ChatGPT agent. https://openai.com/index/introducing-chatgpt-agent/.

Chegg. 2025. Get Homework Help with Chegg Study. https://www.chegg.com/study.

Clarke, R.; and Lancaster, T. 2006. Eliminating the successor to plagiarism? Identifying the usage of contract cheating sites. In *JISC 2nd International Plagiarism Conference*.

Claude. 2025. Claude AI | Your ideas, amplified. https://claude.ai/login?returnTo=%2F%3F.

Dixon, Z.; and George, K. 2021. Monitoring Uncharted Communities of Crowdsourced Plagiarism. *Journal of Academic Ethics*.

Dixon, Z.; George, K.; and Carr, T. 2021. Catching lightning in a bottle: Surveying plagiarism futures. *Online Learning Journal (OLJ)*.

Docsity. 2024. Notes & Exams for university and high school students | Docsity US. https://www.docsity.com/en/usa/.

DocSlides. 2025. DocSlides. https://www.docslides.com/browse.

Dyer, K. A. 2010. Challenges of Maintaining Academic Integrity in an Age of Collaboration, Sharing and Social Networking. In *Proc. of TCC*.

Eaton, S. E.; Chibry, N.; Toye, M. A.; and Rossi, S. 2019. Interinstitutional perspectives on contract cheating: a qualitative narrative exploration from Canada. *International Journal for Educational Integrity*.

Farnham, S.; Turski, A.; and Halai, S. 2021. Docs.com: Social File Sharing in Facebook. In *Proc. of ICWSM*.

Fiesler, C. 2020. Lawful Users: Copyright Circumvention and Legal Constraints on Technology Use. In *Proc. of CHI*.

Fiesler, C.; Feuston, J. L.; and Bruckman, A. S. 2015. Understanding Copyright Law in Online Creative Communities. In *Proc. of CSCW*.

FORCE11. 2020. The FAIR Data principles. https://force11.org/info/the-fair-data-principles/.

Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Iii, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.

Gillis, R. 2019. "Caring about Sharing": Copyright and Student Academic Integrity in the University Learning Management System. In *Copyright Conversations: Rights Literacy: Rights Literacy in a Digital World*.

Hayashi, C. A.; and Fisher-Adams, G. 2012. Who Holds the Copyright in Your Classroom Handout? A Look at Higher Education Work-for-Hire Policies and Students Selling Class Notes. *NCPEA Education Leadership Review*.

Hero, C. 2025a. Course Hero - Copyright Policy. http://www.coursehero.com/copyright/.

Hero, C. 2025b. Course Hero - Service Terms. http://www.coursehero.com/previous-terms-of-use/.

Hero, C. 2025c. Course Hero Plans. http://www.coursehero.com/plans/.

Hero, C. 2025d. Course Hero Summit FAQ - Educators. https://events.coursehero.com/summit-2024/page/2990414.

Holmes, G. L.; and Levin, D. A. 2000. Who Owns Course Materials Prepared by a Teacher or Professor? The Application of Copyright Law to Teaching Materials in the Internet Age. *Brigham Young University Education and Law Journal*.

Ibosiola, D.; Castro, I.; Stringhini, G.; Uhlig, S.; and Tyson, G. 2019. Who Watches the Watchmen: Exploring Complaints on the Web. In *Proc. of WWW*.

Kalker, T.; Epema, D.; Hartel, P.; Lagendijk, R.; and Van Steen, M. 2004. Music2Share - copyright-compliant music sharing in P2P systems. *Proc. of the IEEE*.

Kash, I. A.; Lai, J. K.; Zhang, H.; and Zohar, A. 2012. Economics of BitTorrent communities. In *Proc. of WWW*.

Kolowich, S. 2009. Course Hero or Course Villain?

Kumar, R.; Eaton, S. E.; Mindzak, M.; and Morrison, R. 2020. *Academic Integrity and Artificial Intelligence: An Overview*.

Lederman, D. 2020. Course Hero Woos Professors.

Lieneck, C.; and Esparza, S. 2018. Collaboration or Collusion? The New Era of Commercial Online Resources for Students in the Digital Age: an Opinion Piece. *Internet Journal of Allied Health Sciences and Practice*.

Lou, X.; and Hwang, K. 2009. Collusive Piracy Prevention in P2P Content Delivery Networks. *IEEE Transactions on Computers*.

Lowry, J. 2020. Chapter 4: Ethical Issues in Academic Resource Sharing. *Open Education Alberta*.

L.P., U. N. . W. R. 2022. Best Computer Science Programs - US News Rankings.

Ltd., D. M. C. A. S. 2024a. The Digital Millennium Copyright Act | U.S. Copyright Office.

Ltd., D. M. C. A. S. 2024b. What is a DMCA Takedown?

Mascarenhas, N. 2020. Course Hero, a profitable edtech unicorn, raises rare cash | TechCrunch.

McKenzie, L. 2018. Learning Tool or Cheating Aid?

Meça, A.; and Shkëlzeni, N. 2024. Academic Integrity in the Face of Generative Language Models. In *Emerging Technologies in Computing*.

Mislove, A.; and Wilson, C. 2020. A Practitioner's Guide to Ethical Web Data Collection. In *The Oxford Handbook of Networked Communication*.

Muthukadan, B. 2025. Selenium with Python — Selenium Python Bindings 2 documentation. Licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Office, U. C. 2025. What is Copyright?

Office, U. S. C. 2020. Copyright Registration for Derivative Works.

Peiser, J. 2022. A professor found his exam questions posted online. He's suing the students responsible for copyright infringement.

Perkins, M. 2023. Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*.

Quizlet. 2025. Flashcards, learning tools, and textbook solutions. https://quizlet.com/.

RocketReach. 2025. Docsity Information.

Rodriguez, J. E.; Greer, K.; and Shipman, B. 2014. Copyright and You: Copyright Instruction for College Students in the Digital Age. *The Journal of Academic Librarianship*.

Rodriguez, P.; Tan, S.-M.; and Gkantsidis, C. 2006. On the feasibility of commercial, legal P2P content distribution. *ACM SIGCOMM Computer Communication Review*.

Ruffin, M.; Yan, T.-B.; Adebayo, M.; Wang, G.; and Levchenko, K. 2026. The supplementary materials of this paper. https://tinyurl.com/3vjvs78h.

Seng, D. 2014. The State of the Discordant Union: An Empirical Analysis of DMCA Takedown Notices. *SSRN Electronic Journal*.

Shami, N. S.; Muller, M.; and Millen, D. 2011. Browse and discover: social file sharing in the enterprise. In *Proc. of CSCW*.

Slade, C. 2021. Students as Partners Research Project: Characteristic features of 'buy, sell or trade' file-sharing websites.

Somers, R.; Cunningham, S.; Dart, S.; Thomson, S.; Chua, C.; and Pickering, E. 2024. AssignmentWatch: An Automated Detection and Alert Tool for Reducing Academic Misconduct Associated With File-Sharing Websites. *IEEE Transactions on Learning Technologies*.

Strauss, A.; and Corbin, J. M. 1997. *Grounded theory in practice.* Sage Publications, Inc.

Studocu. 2023. How does Studocu handle copyright issues? https://help.studocu.com/hc/en-us/articles/360047260632-How-does-Studocu-handle-copyright-issues.

Studocu. 2025. Studocu - Free Study Notes for University & High School. https://www.studocu.com/en-us.

StudyLib. 2025. Essays, homework help, flashcards, research papers, book reports, and others. https://studylib.net/catalog/.

Van Wynsberghe, A.; and Van Der Ham, J. 2015. Ethical considerations of using information obtained from online file sharing sites: The case of the piratebay. *Journal of Information, Communication and Ethics in Society*.

Wilson, C.; Boe, B.; Sala, A.; Puttaswamy, K. P.; and Zhao, B. Y. 2009. User interactions in social networks and their implications. In *Proc. of EUROSYS*.

Zhang, W. 2015. Study of early warning of copyright infringement risks in information resources. In *Proc. of DINWC*.

Paper Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes
   (e) Did you describe the limitations of your work? Yes
   (f) Did you discuss any potential negative societal impacts of your work? Yes
   (g) Did you discuss any potential misuse of your work? No, because the potential risk of misuse is minimal.
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? NA
   (b) Have you provided justifications for all theoretical results? NA
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA
   (e) Did you address potential biases or limitations in your theoretical framework? NA
   (f) Have you related your theoretical results to the existing literature in social science? NA
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? NA
   (b) Did you include complete proofs of all theoretical results? NA
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? NA
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? NA
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? NA
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? NA
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? NA
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
   (a) If your work uses existing assets, did you cite the creators? Yes
   (b) Did you mention the license of the assets? Yes
   (c) Did you include any new assets in the supplemental material or as a URL? NA
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? Yes, in Section 3.3 and Appendix A.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes
   (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see FORCE11 (2020))? NA
   (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see Gebru et al. (2021))? NA
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
   (a) Did you include the full text of instructions given to participants and screenshots? No, due to the space limit, we summarized key instructions in the main paper. The full details (including screenshots) were presented in the Appendix and online supplementary materials (hosted under an anonymous link).
   (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? Yes
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? No, our surveys were unpaid, and the participation was voluntary. Further details were in the paper.
   (d) Did you discuss how data is stored, shared, and deidentified? Yes

A Broader Impact and Ethical Statement

Our studies were reviewed and approved by our Institutional Review Board (IRB). Informed consent was obtained from participants at the start of each survey. We did not collect demographic information to keep responses anonymous. All data was securely stored, and student PII was redacted for privacy. Both surveys were unpaid: participants joined the surveys voluntarily and were given the option to opt out or request data removal at any time. While scraping Course Hero content is discouraged by their Terms of Service (Hero 2025b), this method was essential for gathering the volume of data needed for our research. Previous studies have also used web scraping to investigate potentially abusive behaviors on target websites (Aziz and Wilson 2024; Wilson et al. 2009). We prioritized privacy by filtering personal and sensitive information and also maintained a slow crawling rate to minimize the impact on website servers (Mislove and Wilson 2020). Our study seeks to understand the educators' perception of crowdsourced academic document-sharing websites (Course Hero in particular). We emphasize that these results, Course Hero's growth, educators' knowledge and use of DMCA, and more could not have been obtained without this research.

Document Category	Document Count (%)		Total Views (Avg. Views per Doc)	
	w/ Unfair Advantage	w/o Unfair Advantage	w/ Unfair Advantage	w/o Unfair Advantage
Assignment	26 (32%)	**55 (68%)**	**10,376 (399)**	18,427 (335)
Code	**17 (68%)**	8 (32%)	**7,949 (468)**	1,083 (135)
Exam Materials	23 (47%)	**26 (53%)**	**5,527 (240)**	5,810 (223)
Lab	6 (14%)	**30 (83%)**	**3,834 (639)**	18,152 (605)
Quiz	1 (25%)	**3 (75%)**	**1,902 (1,902)**	508 (169)
Review	3 (13%)	**20 (87%)**	**173 (58)**	8,711 (436)
Syllabus	1 (8%)	**11 (92%)**	**106 (106)**	2,111 (192)
Professor Notes	**1 (100%)**	0 (0%)	**15 (15)**	0 (0)
Lecture Materials	1 (8%)	**11 (92%)**	4 (4)	**6,897 (627)**

Table 2: A comparison of documents that give students an unfair advantage to those that do not give an unfair advantage. We break down the comparison of document count and views for each document category. Categories that do not have any "unfair-advantage" documents are omitted from this analysis.

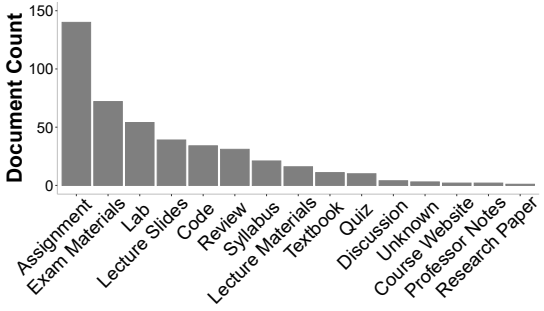

Figure 9: Categories of documents used for the Rights and Integrity Survey (440).

B Rights and Integrity Survey

B.1 Document Categories

When processing documents from Course Hero, we found their labels were often inaccurate. As such, we labeled the documents used in our user study manually (as described in Section 4). Figure 9 shows the categories of documents labeled by us.

B.2 Survey Questions

As discussed in Section 3.2, participants first answered general questions about course practices and their knowledge of DMCA (see Table 3). They were then asked to review five selected course documents and respond to questions Qa1-Qa5. For the final open-ended question, *"Is there anything you'd like to share about the questions or the document-sharing ecosystem?"* we developed a codebook to categorize responses (see Table 4).

Qa1. Take a look at the document above and tell us about its authorship:

- ○ I am the sole author
- ○ Original author but others have made contributions
- ○ Not original author, but I have made substantive contributions
- ○ I am not an author
- ○ I'm not sure
- ○ Prefer not to say

Qa2. If you answered "Not original author, but I have made substantive contributions" to the previous question, please answer the next question. If not, please select Not Applicable. Did the original author place any restrictions on your use of this work (such as not making any changes)?

- ○ Yes
- ○ No
- ○ I'm not sure
- ○ Prefer not to say
- ○ Not Applicable

Qa3. Regardless of the authorship of the document, we believe it is still related to the course you teach/taught. Which of the following best describes the situation?

- ○ I uploaded the document myself
- ○ I authorized someone else upload the document
- ○ I did not authorize the document to be uploaded, but am okay with it being there
- ○ I did not authorize the document to be uploaded, and would like it taken down
- ○ I'm not sure
- ○ Prefer not to say

Qa4. Would this document give an unfair advantage to a student currently taking this class?

- ○ Yes
- ○ No
- ○ I'm not sure
- ○ Prefer not to say

Qa5. To your knowledge, is this document publicly available on the course website, to students not enrolled in the class?

- ○ Yes
- ○ No
- ○ I'm not sure
- ○ Prefer not to say

C Verified Educators Survey

We ask the first question to all participants, which determines if they see any of the following questions. We developed a codebook for our open-ended responses to Qb3 as shown in Table 5.

Qb1. Did you create a faculty account on Course Hero?

- ○ Yes
- ○ No
- ○ I'm not sure

- ○ Prefer not to say

Qb2. Did you upload a profile picture to Course Hero?
(*Shown if answer to Qb1 is "Yes"*)

- ○ Yes
- ○ No
- ○ I'm not sure
- ○ Prefer not to say

Qb3. What was your reason for creating an account? (Select all that apply)
(*Shown if answer to Qb1 is "Yes"*)

- ☐ Take down documents from a course I teach
- ☐ To see if there are documents from a course I teach
- ☐ To access **sample assignments** to use in my own teaching
- ☐ To access **case studies** to use in my own teaching
- ☐ To access **lectures** to use in my own teaching
- ☐ To access **labs** to use in my own teaching
- ☐ To access **syllabi** to use in my own teaching
- ☐ To network with other Faculty
- ☐ Other, please explain

Qb4. Have you used Course Hero to take down documents that were related to your course?
(*Shown if answer to Qb3 is "Take down documents from a course I teach"*)

- ○ Yes
- ○ No
- ○ I'm not sure
- ○ Prefer not to say

Question	Frequency	Percentage
1. Do you have a stated policy in your syllabus and/or on your course website against uploading course content to Course Hero?		
Yes	20	36.36%
No	33	60.00%
I'm not sure	2	3.64%
2. Have you, as an instructor, uploaded any documents to Course Hero?		
Yes	0	0.00%
No	54	98.18%
I'm not sure	1	1.82%
3. To what extent were you aware that documents that referenced your course were being shared on Course Hero?		
I have heard about course documents shared on these websites	21	38.18%
I have checked out the documents myself	11	20.00%
I have filled out a DMCA request / Requested them to be removed	7	12.73%
I have never heard of documents being shared online	12	21.82%
I am not sure if I have heard about them	4	7.27%
4. Do you authorize any material to be shared with others outside of the currently registered students?		
Yes	14	25.45%
No	35	63.64%
I'm not sure	6	10.91%
5. Do you reuse Exam, Quiz, Homework Problems for your course?		
Yes	44	80.00%
No	8	14.55%
I'm not sure	3	5.45%
6. Were you aware that you could request to take these documents down yourself using a Digital Millennium Copyright Act (DMCA) request?		
Yes	18	32.73%
No	31	56.36%
I'm not sure	6	10.91%
7. Have you submitted any DMCA requests yourself?		
Yes	11	20.00%
No	39	70.91%
I'm not sure	5	9.09%
8. What prevented you from doing this yourself? (Select all that apply)		
Too many documents	3	42.86%
Takes too much time	3	42.86%
I do not care that they are up there	3	42.86%
I simply did not want to do it	0	0.00%
9. How much effort are you willing to expend taking down the referenced documents as a whole? (Select all that apply)		
I am willing to fill out a short form (less than a minute) per document	10	18.18%
I am willing to expend one click per document	24	43.64%
I am willing to fill out a short form (less than 5 minutes) for all documents	34	61.82%
I am willing to expend one click for all documents	20	36.36%
I am not willing to do any of the above	7	12.73%
10. If such a tool existed that would automatically submit a DMCA request on your behalf for documents related to this course, would you use it?		
Yes	37	62.27%
No	8	14.55%
I'm not sure	10	18.18%

Table 3: **Rights and Integrity Survey: General Questions** — We conservatively coded PNTS ("prefer not to say") as No.

Primary Code	Subcode	Freq.	Description
Course Hero Content	Solutions matter	10	Documents with solutions on Course Hero are important
Documents Shown	Outdated docs	5	Documents shown are from several years ago
	Mixed authorship	3	Documents shown have many authors who have contributed to them
	Do not adequately represent the class	2	Documents shown do not belong to the course or are from several years ago
	Non-authorship	1	Participant is not an author of any of the presented documents
	Many instructors for course	2	Participant expressed it may be best to ask another instructor of the course as there are many
Negative Sentiment	Negativity towards academic doc sharing sites	9	Expressed great concern for the sustainability of teaching with document sharing sites in existence
	Student unauthorized sharing of material	4	Have found students consistently share course content on Course Hero
	CH unauthorized use of copyright materials	4	Course Hero consistently hosts copyrighted materials
	Students gain unfair advantage from CH materials	3	Course Hero content allows students to cheat and gain an unfair advantage
	Unaware of docs posted online	3	Did not know course content was being posted to Course Hero
	Pessimistic on sharing phenomenon	2	Does not express a good outlook on curbing the unauthorized sharing of documents
	CH material harms students	1	Documents found on Course Hero harms a students chance at learning material
Neutral Sentiment	Docs w/o solutions are okay on CH	4	Documents that do not have solutions may remain on Course Hero
	Indifference toward CH	1	Participant does not care that course content is being shared on Course Hero
	Docs uploaded are not harmful	2	The documents presented were not deemed harmful to students
	Trust students not to cheat	1	Participant wants to trust that students will not cheat
Active Approach	Submit DMCA requests before	2	Participant has submitted DMCA takedown requests before
	Active approach on combating online available solutions	3	Taking steps to reduce the spread of solutions online (i.e., solution honeypots)
Alternative Sharing Methods	Exist other means of circulating materials	8	Course Content is being shared by other means and on other platforms (i.e., GitHub)
	Documents available online	6	Course Content is made publicly available on Course Website
Course Policy	Reuse problems in the course	3	Participants often reuse problems in their assignments
Other Sentiments	Potential research collaboration	2	Participants are doing similar research
	Look forward to study results	1	Eager to see what comes of the study
	Welcome DMCA tool	2	Would use a DMCA tool to take down documents if one existed

Table 4: **Rights and Integrity Survey Codebook: Primary Codes** — We show the code frequencies and related descriptions for our primary codes.

Primary Code	Subcode	Freq.	Description
Check for Illegitimate Use	Check/prove for cheating	5	Use profile to check for assignments and answers
	Copyright check and takedown	3	Use profile to check for IP on Course Hero and issue DMCA requests
Instructional Use	Obtain syllabi for course transfers	3	Cross references syllabi for transfer credit
	Get content for learning	2	Obtain content to use in homeschooling
	Course Hero conference	2	Made profile to attend Course Hero conference
Other Use	Curious about Course Hero	2	Curious about structure of Course Hero and its offerings
	Required by someone else	1	Made to create an account by administration at School
	Can't remember	2	Cannot recall making an account since they do not use Course Hero
	Didn't specify reason	1	No explanation on why an account was made

Table 5: **Verified Educators Survey Codebook: Primary Codes** — We show the code frequencies and related descriptions for our primary codes.