Catching lightning in a bottle: Surveying plagiarism futures

Zachary Dixon
Kelly Whealan George
Tyler Carr
Embry-Riddle Aeronautical University

Abstract

The digitization of higher education is evolving academic misconduct, posing both new challenges to and opportunities for academic integrity and its research. The digital evidence inherent to online-based academic misconduct produces new avenues of replicable, aggregate, and data-driven (RAD) research not previously available. In a digital mutation of the misuse of unoriginal material, students are increasingly leveraging online learning platforms like CourseHero.com to exchange completed coursework. This study leverages a novel dataset recorded by the upload of academic materials on CourseHero.com to measure how at-risk sample courses are to potential academic misconduct. This study's survey of exchanged coursework reveals that students are sharing a significant amount of academic material online that poses a direct danger to their courses' academic integrity. This study's approach to observing what academic material students are sharing online demonstrates a novel means of leveraging digitized academic misconduct to develop valuable insights for planning the mitigation of academic dishonesty and maintaining course academic integrity.

Keywords: academic misconduct, digitization, distance education

As the internet came to gain an increasing role in higher education, some feared a corresponding rise in plagiarism and other forms of academic misconduct. The internet and its affordances seemed to make academically unethical behavior simply too easy for students to resist (Scanlon, 2003). While the more recent scholarship on the frequency and prevalence of plagiarism continues to affirm academic misconduct like plagiarism as a common, important issue in higher education, that research also shows plagiarism is not clearly made worse as a result of the internet (Hart & Morgan, 2010; Ison, 2014, 2015; Kidwell & Kent, 2008; Peterson, 2019; Stuber-McEwen et al., 2009).

Nevertheless, the internet and digitization of higher education do sustain academic misconduct and have enabled new waves of ethically troubling behaviors that challenge academic integrity. Beyond the threat of instant and unacknowledged replication, the internet now supports international contract cheating (Lancaster & Clarke, 2016), peer-to-peer trading of coursework (Rogerson & Basanta, 2016), and the automated production of entire assignments (Shahid et al., 2017). The intersections of higher education and the internet are also opening new methodological opportunities for the analysis of academic misconduct in higher education. The online exchange of academic material, for example, creates new opportunities for research in the digital traces and data that can be used to analyze how, what, and to what degree students share their work. This study leveraged metadata generated from coursework being shared online to develop a novel approach for observing the prevalence of academically dishonest behavior online.

By surveying the kind of materials being shared on CourseHero.com from a sampling of undergraduate courses, this study created a cross-section of how compromised the courses are for academic misconduct. Recognizing the complexities associated with measuring "plagiarism" or "academic dishonesty," this study focused on observing the exchange of unoriginal work as a fundamental condition for those nuanced concepts: the exchange of unoriginal work. Instead of staking out definitive claims about specific student academic misconduct behaviors like plagiarism, this study tried to catch the lightning of what coursework students are exchanging online in a digital bottle as a means of determining how vulnerable courses are to those behaviors. Observing a cross-section of the academic material students shared online served to gauge the propensity or likelihood of academic misconduct by exposing what, and to what degree, formal assignments circulate among students. Such a cross-section provided valuable insights for planning the mitigation of academic dishonesty and of maintaining course academic integrity.

Prevalence of Plagiarism and Academic Dishonesty

Self-reported research on the prevalence of academic misconduct like plagiarism reflects mixed perceptions of its degree of severity but demonstrates frequent or common issues. For example, Hart and Morgan's (2010) survey of online and residential nursing courses reported "very low levels of cheating and very high standards of academic integrity" (p. 501). Wilkinson's (2009) survey of cheating frequency similarly found "less than half of both staff and students thought that cheating in assessment tasks was common" (p. 100). Yet more than a decade ago Scanlon and Neumann (2002) found, "24.5% of . . . students reported plagiarizing online" (p. 381). A more recent self-reported study on cheating behaviors in an Australian university found 15.3% of respondents reported "buying, trading, or selling notes;" 27.2% reported "providing completed assignments to other students;" 5.78% of respondents reporting they had engaged in "one or

more of the five behaviors classified as 'contract cheating'" (Bretag et al., 2019, p. 6). Some research has also suggested academic instructors underestimate the frequency of academic misconduct (Brimble & Stevenson-Clarke, 2005).

However, contrary to early reports of plagiarism's rise in correlation to the internet (Scanlon, 2003; Scanlon & Neumann, 2002), more recent work has shown online respondents do not report more frequent cheating behaviors (Hart & Morgan, 2010; Stuber-McEwen et al. 2009; Kidwell & Kent, 2008). Respondents in Stuber-McEwen et al.'s (2009) survey comparing residential and distance course students even reported students in traditional classroom environments as more likely to cheat than those online. A large-scale survey of residential and distance students measuring the self-reported occurrence of 17 "cheating behaviors" found distance students reported "Considerably less cheating" than their residential counterparts (Kidwell & Kent, 2008, p. S14). Watson and Sottie's (2010) survey of more than 600 undergraduate and graduate students found nearly identical levels of self-reported cheating or academically dishonest behavior between online and residential classes. The same study also found "for almost every individual survey statement, more students admitted to inappropriate behavior in face-to-face classes than in online courses" (Watson & Sottie, 2010, p. 5). More empirical measures of plagiarism's frequency, typified by the use of text-matching or "similarity detection" software like Turnitin or SafeAssign, paint a more detailed, but still varied image. The application of text similarity analysis on student coursework generally demonstrates the widespread occurrence of problematically similar or even exact, unattributed text in student writing which is a common benchmark for plagiarism across higher education (Table 1).

Table 1

Results of Similarity Detection Plagiarism and Their Methods of Measure

Study citation	% of samples demonstrating "plagiarism"	Analysis tool/method (qualification threshold)
Waker (2010)	26.2%	Turnitin "similarity index" and custom analysis (not specified)
Martin et al. (2011)	61%	Turnitin similarity index (3%)
Gilmore et al. (2010)	42.6%	SafeAssign index (not specified)
Chao et al. (2009)	39%	Turnitin similarity index (at least 1%)
Ison & Szathmary (2016)	39.6%	SafeAssign index (15%)
Ison (2012)	72%	Turnitin "similarity index" (10%)
Martin et al. (2009)	61%	Turnitin "similarity index" (3%)
Batane (2010)	66.1%	Turnitin "similarity index" and custom analysis (1-34%)

Ison (2018)	25.1%	Turnitin "similarity index" (not specified)
Holmberg & McCullough (2006)	63.2%	Custom analysis
Ison (2015)	77% (pre-internet), 52.8% (post-internet)	Turnitin "similarity index" (at least 11%)

Ison's (2012) study of dissertations from a distance Ph.D. program found 72% of samples had "at least one case of improper paraphrasing and citation" and 46% of samples had "verbatim text without citation" (p. 233). On aggregate, 46% of sampled dissertations "were classified as having a low level of plagiarism," 11% with a "medium level," and 3% with a "high level" (Ison, 2012, p. 233). In another similar detection study, 40% of sample "capstone assignments" from a cohort of graduate student capstone courses exceeded a SafeAssign index threshold of 15% (Ison & Szathmary, 2016). Measuring the Turnitin "similarity" index score of dissertations from various global regions showed the improper use and/or attribution of unoriginal material is a common issue, with samples demonstrating a mean similarity index of 25.1% (Ison, 2018); this means, on average, a quarter of the writing from sampled dissertations was flagged as problematically similar to unattributed sources. The same study also found little statistically significant variance between global regions defying "the assumptions of rampant plagiarism and other forms of academic misconduct in specific countries and regions" (p. 302). Walker (2010) found the unattributed use of unoriginal material to be a relatively frequent occurrence, with more than a quarter (26.2%) of sampled student work demonstrating "some sort" (p. 48) of plagiarism. In those samples, unattributed paraphrasing was the most common manifestation of plagiarism (15.7%) and the substantial or entire submission of unoriginal work was the least common (1%) (Walker, 2010). Similarly, although Ison's (2015) findings show a majority of pre- and post-internet dissertations contain problematic text, the plagiarism was generally "low level" with a "similarity index range" of 11 to 24%.

The use of similarity detection analysis further erodes the correlation between distance education, the internet, and plagiarism's prevalence. Peterson's (2019) review of research on the differences in academic dishonesty between online and residential classes shows "little evidence that cheating is more prevalent in online courses" (p. 33). Ison's (2014) comparison of Turnitin "similarity indexes" on dissertations from residential and online institutions found "no statistically significant difference in the level of plagiarism" (p. 278) between types of institutions. Comparing dissertations submitted before the widespread use of the internet in academia (1991–993) against more current dissertations (2010–2014) even showed pre-internet work to have higher mean Turnitin "similarity index" scores, thus contradicting the notion that the digital environment has increased the misuse of others' work (Ison, 2015).

Plagiarism Futures
Still, the internet clearly supports academically dishonest behavior, and that digitized behavior is happening in new spaces and manifesting in new ways. Academically dishonest behavior is evolving in digital marketplaces that facilitate contract cheating (Lancaster & Clarke, 2016), peer-to-peer (Rogerson & Basanta 2016) or crowd-sourced sharing of coursework (Dixon

& Whealan George, 2020), and even terrifyingly futuristic spun manuscripts (authored entirely by machine learning or artificial intelligence software) that evade detection by normal text-matching software (Shahid et al., 2017). Contract cheating is facilitated by worldwide digital markets that connect students to for-purchase, original, third-party coursework that is extremely difficult to catch and good enough to pass assessment (Medway et al., 2018; Malesky et al., 2016). Online contract cheating markets are reflective of an increasingly transactional approach to education where students view it "as a product to be bought, sold or traded rather than an intrinsically motivated, effortful and potentially transformative individual process" (Bretag et al., 2019, p. 2). This transactional view of education is sustained by web platforms explicitly designed for the exchange of academic coursework. An increasing number of web platforms like Chegg, Coursehero, Quizlet, and Study.com offer students the infrastructure to exchange coursework in peer-to-peer (Rogerson & Basanta, 2016) or crowd-sourced (Dixon & Whealan George, 2020) fashion. Coursework exchange web platforms capitalize on nuances of authorship and ownership and blur the line between scholarly collaboration and academic dishonesty by facilitating the "strong temptation for students to reuse or repurpose downloaded content for personal gain and academic advantage" (Rogerson & Basanta, 2016, p. 265) without proper attribution. As Rogerson and Basanta (2016) further argued,

> *There is a big difference between sharing knowledge based on the principles of academic integrity versus information uploading and downloading under the guise of supporting others, which ultimately conceals or obscures original authorship and potentially distorts content and meaning. (p. 265)*

The online environment facilitates complex forms of information exchange and reproduction that are difficult to define, detect, or even observe. Coursework exchange platforms and the academic behavior they support are obscured by logins, passwords, proprietary user-agreements, and obstructive community standards (Dixon & Whealan George, 2020). Even though these digital evolutions of academic dishonesty challenge higher education to reflect on "the ways in which the sharing economy is shaping students' approaches to life and learning" (Bretag et al., 2019, p. 22), there has been little evidence of or investigation into how these coursework exchange platforms are used by students.

Despite the challenges that online academically dishonest behavior may pose to higher education, however, digitalization may also offer a methodological boon to academic misconduct researchers. The exchange of coursework—arguably a cornerstone of academically dishonest behaviors usually considered plagiarism—now comes with digital traces. Whereas exchanging coursework once depended on largely private and hidden interactions, that exchange is now recorded with digital evidence. For example, coursework uploaded to exchange platforms include background metadata describing document characteristics like when it was uploaded, which user uploaded it, and what its primary content is about. Digital evidence like document metadata opens new avenues of replicable, aggregate, and data-driven (RAD) research not hitherto available for academic misconduct research. Haswell (2005) defines RAD research as "a best effort inquiry" that is "explicitly enough systematized in sampling, execution, and analysis to be replicated; exactly enough circumscribed to be extended; and factually enough supported to be verified" (p. 201). RAD research describes a quantitative, usually computer-enabled methodological approach to research topics normally that are typically analyzed from a qualitative perspective.

The digitization of academic misconduct and the potential for RAD academic integrity research is analogous to that of writing and composition in higher education more generally, shortly after the turn of the millennia. As student writing increasingly took place in digital formats rather than pen and paper, their work compiled into corpuses of data. Instead of stacks of physical papers, students increasingly generated bytes of data. Digital corpuses enabled Composition Studies researchers to apply computer-assisted analysis methods like concordance software, which can measure word frequencies and patterns, among other aspects to re-test pedagogy and assessment research with more RAD methodologies. The digital migration of student writing made possible research that would otherwise be difficult or even unworkable (Fishman, 2012; Haswell, 2012; Dixon & Moxley, 2013). Since digital writing corpuses represent more stable, finite, shareable data, analysis of that data is more exact, systematic, replicable, and verifiable. As Dixon & Moxley (2013) note of their study of more than 100,000 instructor comments on student writing, digitization of the corpus and analysis enabled "in a few keystrokes what once took years" (Dixon & Moxley, 2013, p. 243). By enabling more RAD methodologies, the digitization of composition facilitated an "increased sensitivity to the local contexts, rhetoric, and characteristics of writing" (Dixon & Moxley, 2013, p. 252). Such sensitivities contribute to research that meaningfully captures the subjects, purposes, and meanings of writing (Dixon & Moxley, 2013). The migration of plagiarism and academic misconduct into online spheres is creating similar digitized research opportunities. The academically dishonest exchange of coursework and other academic materials is now recorded in timestamps, texts, emails, IP addresses, uploads, downloads, and metadata. Plagiarism and academic integrity or misconduct research now have datasets primed for more RAD research.

One such dataset is the academic material and coursework being shared on CourseHero.com. CourseHero.com is an online learning platform offering course-specific study materials from over "40 million course materials" (2019a, para. 1). In addition to a catalogue of tutoring and Q&A services, textbook resources, and other study materials, CourseHero.com hosts the exchange of syllabi, questions, instructor notes, homework solutions, complete essays, completed tests, and other coursework produced by students. Students either pay for access to CoureHero's database or can upload 10 documents to "unlock" 5 downloads (CourseHero.com, 2019b). While the corpus of student work hosted by CourseHero.com alone does not clearly constitute academic misconduct, it does embody a broad transition zone between social learning and academic dishonesty. CourseHero.com's vast trove—collected with the intent of exchange by students—establishes the ideal conditions for academic dishonesty through the use and submission of unattributed, unoriginal academic material as students' own. In this way, CourseHero's digital corpus of academic materials represents a kind of plagiarism futures trading. Futures trading essentially contracts another party to pay for an asset today, with delivery at a future date, at a predetermined price. To relate this to the plagiarism, students' participation in CourseHero.com essentially entails agreeing to pay for an asset via download, at a future time by the student, for a predetermined price for a set number of uploads.

Methods

Research Design

This study focused on the academic materials shared on CourseHero.com from a sampling of courses to develop an image of one university's plagiarism futures. This study used a descriptive research design to survey what and how much coursework is being shared by students online. This study's research design did not begin with a hypothesis, but sought a measure of how compromised, or how at-risk for compromise, a course is as a result of its assignments and assessments being available to potential misuse.

Setting

The sample university was a mid-sized private institution (30,000 students), supporting undergraduate and graduate degree programs across two residential campuses, and a distance education campus. Researchers selected a group of eight undergraduate courses from the distance campus catalogue that were frequently taught. Multiple sections of the selected courses are offered every term and are typically full to enrollment capacity. The selection of undergraduate courses ensures almost all of the University's student are likely take one of the sampled courses, thus assuring a thorough sample population of students. This study did not require IRB approval because it gathered data that was publicly available, did not require the researchers to observe, interact, or intervene with individuals to gather the data, nor did any of its analysis, results, or conclusions utilize any personal identification data. This study monitored the coursework being shared on CourseHero.com from the sample university's following course prefixes: HUMN330 Values and Ethics; WEAX201 Meteorology I; ENGL123 English Composition; ECON211Macroeconomics; RSCH202 Introduction to Research Methods; MATH11 College Math for Aviation; UNIV101 College Success; PHYS102 Explorations in Physics. This selection of courses ensured a variety of sample disciplines and a constant stream of active students taking the courses during the Spring 2020 term to best simulate the normal changes in artifacts found on the website.

Data Collection

Data collection was facilitated by a custom application titled Course Villain. Course Villain is an original, web-based desktop application developed by a faculty and student research team at the sample university. Course Villain was designed with the explicit purpose of monitoring the uploading of university content on CourseHero.com by performing automatic, custom searches, aggregating results matching search terms, and engaging CourseHero.com's "Copyright Infringement" workflow to remove content matching query terms. Course Villain used a webserver that performed scans and ran a database of results, and a desktop application for the user interface that displayed results with a browser that allowed for automation. The Course Villain software was experimental in design, ongoing in development, and has not been subjected to rigorous reliability or usability testing.

Course Villain users, which at the time were faculty and student researchers, start by downloading the desktop application for either Mac or Windows, and then create an account. Once users have an account, they define search queries for specific courses that they want to monitor. Query parameters filter results to search for the course name, ensured all document types were shown, and constrain results to the sample university's content on CourseHero.com. Scans for new documents uploaded to CourseHero.com for all courses were performed twice per day by the software. New query matches were recorded to a database for users to view and were also sent to users through email reports. Users can view documents matching query terms through a page in the desktop application. Users can choose to either ignore documents if they are irrelevant or have the application automatically populate the CourseHero.com "Copyright Infringement" form with information about the researcher, the document, and the course the

document belongs to. Populating the "Copyright Infringement" form must be performed on the user's computer instead of a website because the CourseHero.com "Copyright Infringement" form contains a Google reCAPTCHA that requires a user to complete a task proving they are not a robot in order for the form to be submitted. Documents that have been marked as irrelevant or have already been reported are labeled as such in the database, are hidden from the interface. If a document that has been reported is taken down from CourseHero.com, the reporting user will receive an email correspondence from them.

Course Villain scans are run on the webserver using a headless, or invisible, browser window. For all course queries created by users, CourseHero.com opens new tabs and searches for documents belonging to this course. Course Villain's search windows are generated and controlled using Puppeteer, which is automation software created by Google to control browsers using code. Search pages are filtered to show only content from the sample university. All documents on search pages are scanned and each document title or name is compared against the search query terms to ensure relevancy. Information about each document matching query terms is taken directly from the search page and saved to the database. Document names, IDs, upload dates, and document types are all recorded by Course Villain. If a document is already recorded in the database, it will be skipped. To minimize false-positive matches, the course title or name metadata of each document is compared against the query terms by the software. A document is only recorded with exact matches. Document course name and query term comparison also improves scan accuracy by locating artifacts whose other metadata may not accurately match query terms or criteria The first time a query is scanned, the search is organized by relevance according to most recently uploaded to CourseHero.com, and all documents with matching course names are recorded for each page until there are no documents on a page. Initial scan results are organized by recency and all pages are scanned until the last page is reached. This routine maximizes the number of documents recorded during the scans. Subsequent query searches only the artifact first page, using the recency filter to control for newly uploaded documents.

Course Villain uses the Node.js runtime environment. Node.js was created in 2009 and has been trusted and implemented by many major technology companies since then (Brewster, 2020). Course Villain's functional reliability over time depends mostly on revisions to the CourseHero.com application program interface (API). From Course Villain's initial development in 2018 to its most recent update in 2020, the CourseHero.com website changed multiple aspects of its design, including both the search result pages and the "Copyright Infringement Notification" form. These revisions to CourseHero.com made the Course Villain application unusable and required coding changes to function. Whenever CourseHero.com changes, Course Villain's code must be revised accordingly.

For this study, search queries for the aforementioned course prefixes were run for one academic term (one nine-week academic term at the sample university). Course Villain collected materials already uploaded on CourseHero.com prior to the beginning of the scan, as well as newly uploaded material during the test period. To evaluate the artifacts collected from CourseHero.com, the researchers manually categorized the artifacts into low, medium, and high value categories. The categorization of artifacts was based on the researchers' interpretation of how an artifact might jeopardize or endanger a course's assessments if publicly available. The researcher's categorization was based on an artifact's point value or weight in terms of their course's assessments or grades and on a subjective qualification of how severe an impact the misuse of a given artifact would have on the course's academic rigor and integrity. For example,

the degree to which a student's personal notes about a lecture would compromise the integrity or rigor of a course is much different than a final exam's answer set. The categories created were as follows:

Low: notes, syllabus, PowerPoint presentations

Medium: homework, discussion questions, problem sets

High: quizzes, tests, papers, case studies

Other: artifacts that are not related to course

Researchers then calculated a "compromise metric" using the categorized data. The compromise metric is equal to the sum of medium and high value assignments, divided by the total number of artifacts recovered from a given course. A compromise metric near or greater than 50% was considered alarming, whereby a course could be considered significantly compromised. For example, if an instructor planning for an upcoming term was presented with the information that half of the course deliverables were already available on the internet, that instructor would likely implement significant revisions to the course's assignments. The compromise metric intentionally underweights the percentage of a course that is compromised because the low category like notes, syllabus, and PowerPoint presentations are typically not included in students' final grades.

Data Analysis

Data analysis was completed using a simple spreadsheet software. The faculty researchers manually opened each document flagged by Course Villain and visually reviewed the artifacts. Faculty researchers then categorized the artifact into the appropriate predetermined categories of low, medium, high, and other. In many cases, it was unnecessary to manually open the artifact because the title assigned to the document uploaded by students was evident to which category they should be tabulated within. For example, titles like Final Paper, Final Exam, Module 3 Discussion Questions, or Module 5 Problem Set allow quick assignment without the more time-consuming task of opening the artifact and viewing the submission. Once all collected artifacts were categorized, the researcher calculated the percentage of assignments collected in each category and calculated the compromise metric for each course in the sample.

Limitations

The specific results of the study offer little external validity. Instead, this study's design is intended to produce detailed, internally valid results that render an image of the sample university's specific contexts. Without purposefully and accurately capturing the context of other Universities or institutions, this study's results do not and cannot provide meaningful, specific conclusions about academic misconduct or plagiarism writ large. Instead, this study's results provide the sample university with detailed evidence about its courses' potential for academic misconduct. Future research is planned to use this software for external validity and application on how online coursework sharing might be studied.

Another limitation inherent to this study's design is the dependence on descriptive statistics. Without robust inferential statistics, this study's results cannot offer valid conclusions about patterns or predictions of online coursework sharing, even in the specific context of the sample university. Instead, this study's results offer only a general description of a novel dataset captured from a complex practice.

This study's final significant limitations stemmed from the absence of reliability testing and subjective coding in analysis. As noted above, the Course Villain application's scan results were inherently subject to a measure of reliability error. Course Villain's results were subject to user error in query term design, incorrect matches due to metadata misidentification, and

unaddressed changes to CourseHero.com's API. Through continuing development and maintenance, the research team has worked to mitigate errors, particularly as detailed in the Methods section. False-positives and other kinds of result errors were also screened out in analysis. However, Course Villain has not been subject to rigorous reliability testing. Additionally, this study's results hinged on the researcher's subjective coding of Course Villain's results. The application facilitated scanning, collecting, and organizing the coursework artifacts available to students on CourseHero.com, but analysis required manual classification of the artifacts. The study's results, therefore, were influenced by the researchers' subjective interpretation of sampled documents' value and no inter-rater reliability measures were undertaken during this study. Whenever possible, artifacts were judged according to file names, document titles, or other obvious metadata recorded by Course Villain. When a document was titled "Final Test" or "Final Paper," it was fairly easy to characterize that as a high value artifact. When simple artifact identifiers could not be used for evaluation, researchers reviewed what was available through Coursehero.com's document preview.

Results

The test period was successful in scanning for and collecting artifacts in seven of the eight courses. Over the nine-week academic term, Course Villain produced 92 reports from across the sample courses, capturing 1,890 artifacts. One of the courses, ECON 211, was misidentified in query terms and did not return any results. However, with 13 reports capturing 260 artifacts for each course, more than enough data were collected for the remainder of the courses to adequately survey the types of artifacts present on CourseHero.com, as well as a calculation of the compromise metric (see Table 2).

Table 2

Compromise Metric Results in Terms of Value and Types of Artifacts Collected

Course	Compromise Metric	Low Value	Medium Value	High Value
HUMN 330	100%		Discussion Questions (60%)	Tests, Papers (40%)
WEAX 201	44%	Notes (56%)	Homework, Discussion Questions (44%)	
ENGL 123	31%	Notes, study documents (69%)	Discussion Questions (30%)	Papers (1%)
ECON 211	n/a	n/a	n/a	n/a
RSCH 202	56%	Notes (38%)	Homework, Discussion Questions (41%)	Quizzes, Papers (15%)

MATH 111	71%	Misc. Documents (33)	Homework, Discussion Questions (50%)	Quizzes, Papers (17%)
UNIV 101	0%	Notes (100%)		
PHYS 102	46%	Misc. Documents (54%)	Homework, Discussion Questions (46%)	

Overall, half of the courses in this study demonstrated compromise metrics of nearly 50% (49.7% actual mean value), meaning that almost half of all artifacts collected represent graded deliverables vital to the academic rigor of the courses. All of the course materials shared from HUMN330, Value and Ethics, represented significant threats to its academic rigor or integrity. Of HUMN330's shared materials, 60% were discussion question responses or other kinds of homework exercises, and 40% were completed test answers or whole essay assignments. At the other end of the observed spectrum, none of the material shared from UNIV101 (College Success), posed a meaningful threat to its rigor or integrity. All of the observed materials being shared from UNIV101 were some kind of student notation and did not represent actual, assessed coursework. MATH111 (College Math for Aviation) and RSCH202 (Introduction to Research) both scored high compromise metrics of 71% and 56%, respectively. MATH111 and RSCH202's compromised materials were also both spread across all kinds of assessments. A generally random collection of non-assessed artifacts comprised 33% of MATH111's shared materials, while 41% were homework answer sets or discussion question responses, with 17% being completed quizzes or major paper assignments.

Similarly, 38% of RSCH202's shared materials were student notes, 41% were completed homework assignments or discussion question responses, and 15% were completed quizzes or major paper assignments. PHYS102 (Explorations in Physics) also recorded a high compromise metric of 46%, though that metric was composed primarily of low- and medium-value assessments. Student notes accounted for 54% of PHYS102's shared materials and completed homework assignments or discussion question responses accounted for 46%. ENGL123 (English Composition) recorded a 30% "compromise metric" with student notes accounting for 69%, discussion question responses accounting for 30%, and major papers or essays accounting for 1% of the materials being shared. WEAX201 (Meteorology 1) recorded an overall compromise metric of 44%, with 56% of shared materials representing student notes, and 44% representing discussion question responses or other kinds of homework exercises.

Discussion

Given the relatively limited duration and scope of this study's design and the many nuances inherent in defining and measuring academic misconduct like plagiarism, external and valid conclusions are not appropriate. Within the context of the sample university, however, this study's results demonstrate that surveyed courses are worryingly compromised by the exchange of coursework on CourseHero.com. An aggregate mean compromise metric of 49.7% among the sampled courses shows nearly half of materials shared by students on CourseHero.com was identified as either medium or high value to course integrity. Such exchange likely endangers the value and integrity of those course assessments, and this study's results give strong testimony for

urgent course revision. With all but one of the surveyed courses demonstrating a compromise metric of greater than 30%, it is additionally clear that students from the sample university are exposing a significant degree of coursework that poses a meaningful danger to the academic integrity of those courses.

This study's results show that problematic coursework exchange is slightly more prevalent among sampled STEM subjects than others. Even with Value and Ethics' (HUMN330) 100% compromise metric, sampled non-stem subject courses shared a mean of 43.6%. College Success (UNIV101), a general education course that introduces students to fundamental aspects of being a student in higher education, was the only course to score a zero-compromise metric. The four STEM subject matter courses recorded a mean compromise metric of 54.25%, signaling that more than half of all the coursework shared from the courses by students represents a meaningful danger to the courses' academic integrity. College Math for Aviation (MATH111) scored a 71% compromise metric, showing that most of its exchanged coursework is dangerous to its integrity. Introduction to Research (RSCH202) also recorded a notably high compromise metric of 56% which was a level the researchers considered a direct threat to the course's integrity. The notable difference between the compromise metric of STEM and non-STEM signals a potentially notable finding worth further testing.

The observed coursework exchanged among sampled STEM courses is particularly worrisome because their subject matter content is arguably more objective or finite, and less flexible in how their basic materials might be appropriately used by students than the non-STEM sample. By example, the catalogue of coursework being exchanged from MATH111 is less open to subjective interpretation, reuse, and alteration than that of HUMN330. Arriving at the results of an algebraic equation by virtue of downloading the assignment is more clearly an act of academic dishonesty than downloading another student's current events blog to inspire one's own writing process. In this way, HUMN330's 100% compromise metric is startling and more careful analysis of how those shared materials were used by students is necessary to draw meaningful conclusions about the connection between that exchange and academic misconduct. The 71% of medium- and high-value coursework shared from MATH111 is more problematic because there are fewer conditions in which sharing completed quiz and test question and answer sets is appropriate. These results suggest students at the sample university are not only sharing a meaningful degree of coursework that poses a danger to those courses' academic integrity, but also the coursework being exchanged seems likely or directly connected to academic misconduct.

Though this study's results do not provide a clear or direct measurement of plagiarism, they may still be informative to compare against those from self-reported and similarity-detection studies of plagiarism. A mean compromise metric of 41.3% among the sampled courses may indicate that academic misconduct is slightly more prevalent than reported by survey-based plagiarism research methodologies, but towards the lower end reported by similarity detection methodologies. With detailed self-reported data showing roughly a quarter of students admitting to various kinds of academically dishonest behavior (Bretag et al., 2019), this study's findings exceeded that measure with a mean 41% of the coursework being shared from sampled courses. This study indicated a meaningful threat to the academic integrity of the courses as a higher level of questionable academic behavior is taking place than previous research suggests. The margin between this study's findings and those of self-reported academic misconduct research may be explainable by students' observed lack of understanding about what constitutes academic misconduct (Gullifer & Tyson, 2010; Ramzan, 2012; Hu & Lei, 2015), and

the gaps between student and instructor perspectives about misconduct (Watkinson, 2009; Brimble & Stevenson-Clarke, 2005). If students do not fully understand what academic misconduct is, or what their instructors and institution's expectations about it are, they are not likely to accurately report misconduct behaviors.

With medium- and high-value course artifacts accounting for 31 percent to 100 percent of the coursework being exchanged by students in sampled courses, this study seems to generally reflect the findings of similarity detection methodology research (see Table 1). The rate at which important coursework is being shared supports similarity detection-based findings that plagiarism is, while not an overwhelmingly frequent behavior, nevertheless common and serious. However, these results are likely less congruent than might appear. This study's observations do not completely capture the breadth or depth of plagiarism behaviors since illicitly exchanging coursework online is only one of many other possible means of plagiarizing. This study's results, therefore, likely underrepresent the improper use and attribution of unoriginal material in coursework. This study's observations also cannot account for what students actually do with exchanged coursework, and therefore likely overrepresent the academically dishonest or unethical behavior under scrutiny. It is unreasonable to assume that all of the coursework being exchanged in these spaces is being used in nefarious ways.

The widespread exchange of compromising coursework observed in this study suggests that crowd-sourced plagiarism represents a meaningful issue for the sample university, and perhaps also for similar distance education campuses. While this study's limited design and analysis do not directly contribute to scholarship about the prevalence of academic misconduct in distance education, it does signal the prevalence of ideal conditions for academic misconduct online. In the same way that seasonality makes parts of the equatorial Atlantic Ocean likely to sustain the development of tropical cyclones, the exchange of coursework on CourseHero.com appears to be promoting prime conditions for academic misconduct. Students are clearly exchanging a significant degree of problematic coursework online. For the sample university's campus whose courses were sampled, the increasingly favorable conditions observed here signal potential storms on the horizons of its courses' academic integrity.

Conclusion

True to any forecasting, perfect foresight into the nature and severity of academic integrity's coming storms is impossible. This, however, does not render proactive action against academic misconduct like plagiarism impossible. As Sutherland-Smith (2016) concluded, effectively mitigating academic misconduct requires a pluralized approach, necessitating diverse angles of "dialogic processes, academic research, collegial action, effective policy and reflexive teaching" (p. 40). Rather than cure-alls, the most actionable approaches to combat academic misconduct are found in the careful details of how and what disciplines, institutions, and instructors need and want to help students accomplish.

This study's most important contribution to academic misconduct research is to demonstrate a novel approach to monitoring academic misconduct behaviors. As noted above, the students' exchange of coursework online should not be considered equivalent with academically inappropriate behavior. Rather, keeping scholarly tabs on how much and what coursework students are actually exchanging with one another online is a promising and relevant means of better understanding the practice of students intentionally submitting the work of others as their own which is a keystone of plagiarism and other kinds of misconduct. In addition to self-

reported student and instructor perspectives, and similarity detection software, monitoring the online exchange of coursework offers another vector by which to triangulate academic misconduct.

In addition to self-reporting methodologies, digitized coursework exchange monitoring offers a more RAD approach to academic misconduct, grounded in new data and analysis techniques not previously available. Even somewhat rudimentary programming like the kind used in this study's data collection taps into plagiarism futures' digital records in more aggregate and replicable ways than survey-based approaches. Digitized coursework exchange monitoring also compliments similarity detection approaches to defining and measuring academic misconduct by more explicitly maintaining researchers' vital contexts. Similarity detection ultimately hinges on black-boxed algorithms and proprietary corporate products outside of the researcher's control. Digitized coursework exchange monitoring, rather, depends on academic misconduct researchers to define their own terms of observation and measurement, making clearer their cultural, disciplinary, and pedagogical frames that dictate their perspectives on academic misconduct. Without such context, observations of digitized coursework exchange will lack relevancy, rigor, and application.

Observing the exchange of coursework online also welcomes a valuable measure of currency to both more established approaches to studying academic misconduct. Both survey and similarity detection approaches to academic misconduct research capture problematic behavior after the fact. Capturing and researching academic misconduct after it has been perpetrated contributes to a largely reactive stance. Monitoring what coursework is being exchanged in closer to real-time conditions gives researchers, instructors, and administrators a kind of foresight on which courses are or may be becoming vulnerable to academic misconduct; putting them in a position to be more proactive. This sort of foresight is particularly relevant and valuable in distance education contexts that rely heavily on instructional design. Assessment design and management plays a vital role in mitigating the danger posed by peer-to-peer sharing to academic integrity (Rogerson & Basanta, 2016). However, whereas residential instructors can adjust their curricula to notable classroom trends somewhat quickly, as needed, distance instructors frequently face more lag-time. Asynchronous online courses built around course-shells, for example, may have a refresh or redesign timeline of academic years instead of lectures or weeks. Monitoring the digital exchange of coursework in distance education contexts gives higher education stakeholders a more current means of anticipating which courses, or even particular assignments should take precedence in the queue of revision. In either residential or distance education contexts, digital coursework exchange monitoring provides a more current, proactive means of engaging a holistic academic integrity culture.

Ultimately, careful triangulation is the best approach to addressing academic misconduct in higher education. The work here demonstrates the potential for new prong of academic misconduct research, focused on a new mutation of a classic issue with equally novel methods. Self-reported methodologies give academic misconduct research the means to reveal and decipher faculty and student perspectives. Similarity detection lends academic misconduct research a measure of objectiveness which helps codify and parse misconduct behaviors. Monitoring the digital exchange of coursework offers higher education's researchers, administrators, and instructors an additional, particularly current, and data-driven means of triangulating academic misconduct in their own vital contexts.

References

Batane, T. (2010). Turning to Turnitin to fight plagiarism among university students. *Journal of Educational Technology & Society*, *13*(2), 1–12.

Bretag, T., Harper, R., Burton, M., Ellis, C., Newton, P., Rozenberg, P., . . . & van Haeringen, K. (2019). Contract cheating: A survey of Australian university students. *Studies in Higher Education*, *44*(11), 1837–1856.

Brimble, M., & Stevenson-Clarke, P. (2005). Perceptions of the prevalence and seriousness of academic dishonesty in Australian universities. *The Australian Educational Researcher*, *32*(3), 19–44.

Cheema, Z. A., Mahmood, S. T., Mahmood, A., & Shah, M. A. (2011). Conceptual awareness of research scholars about plagiarism at higher education level: Intellectual property right and patent. *International Journal of Academic Research*, *3*(1), 666–671.

Chao, C. A., Wilhelm, W. J., & Neureuther, B. D. (2009). A study of electronic detection and pedagogical approaches for reducing plagiarism. *The Journal of Research in Business Education*, *51*(1), 31–42.

Course Hero. (2019a). *About us.* Retrieved August 8, 2021, from https://www.coursehero.com/about-us/

Course Hero. (2019b). *Upload*. Retrieved from https://www.coursehero.com/upload/.

Dixon, Z., & Moxley, J. (2013). Everything is illuminated: What big data can tell us about teacher commentary. *Assessing Writing*, 18(4), 241–256.

Dixon, Z., & Whealon George, K. (2020). Uncharted communities of crowdsourced plagiarism. *Journal of Academic Ethics*. https://doi.org/10.1007/s10805-020-09381-2

Eret, E., & Ok, A. (2014). Internet plagiarism in higher education: Tendencies, triggering factors, and reasons among teacher candidates. *Assessment & Evaluation in Higher Education*, *8*(31), 1002–1016.

Fishman, J. (2012). Longitudinal writing research in (and for) the twenty-first century. In G. Kirsh (Ed.), *Writing studies research in practice: Methods and methodologies* (pp. 171–182). Southern Illinois UP.

Gilmore, J., Strickland, D., Timmerman, B., Maher, M., & Feldon, D. (2010). Weeds in the flower garden: An exploration of plagiarism in graduate students' research proposals and its connection to enculturation, ESL, and contextual factors. *International Journal for Educational Integrity*, *6*(1). https://doi.org/10.21913/IJEI.v6i1.673

Gullifer, J., & Tyson, G. A. (2010). Exploring university students' perceptions of plagiarism: A focus group study. *Studies in Higher Education*, *35*(4), 463–481.

Hart, L., & Morgan, L. (2010). Academic integrity in an online registered nurse to baccalaureate in nursing program. *Journal of Continuing Education in Nursing*, *41*(11), 498–505.

Haswell, R. H. (2005). NCTE/CCCC's recent war on scholarship. *Written Communication*, *22*(2), 198–223.

Hayes, N., & Lucas, D. Introna (2005) Cultural values, plagiarism, and fairness: When plagiarism gets in the way of learning, *Ethics & Behavior*, *15*(3), 213–231.

Holcomb, C., & Buell, D. A. (2018). First-year composition as "Big Data": Towards examining student revisions at scale. *Computers and Composition*, *48*, 49–66.

Holmberg, M., & McCullough, M. (2006). Plagiarism in science and technology master's theses: A follow-up study. *New Review of Information Networking*, *12*(1–2), 41–45. doi:10.1080/13614570601136255

Hu, G., & Lei, J. (2015). Chinese university students' perceptions of plagiarism. *Ethics & Behavior*, *25*(3), 233–255.

Ison, D. C. (2018). An empirical analysis of differences in plagiarism among world cultures. *Journal of Higher Education Policy and Management*, *40*(4), 291–304.

Ison, D. (2017). Academic misconduct and the Internet. In D. Velliaris (Ed.) *Handbook of research on academic misconduct in higher education* (pp. 82–111). IGI Global.

Ison, D. C. (2015). The influence of the Internet on plagiarism among doctoral dissertations: An empirical study. *Journal of Academic Ethics*, *13*(2), 151–166.

Ison, D. C. (2014). Does the online environment promote plagiarism? A comparative study of dissertations from brick-and-mortar versus online institutions. *Journal of Online Learning and Teaching*, *10*(2), 272–281.

Ison, D. C. (2012). Plagiarism among dissertations: Prevalence at online institutions. *Journal of Academic Ethics*, *10*(3), 227–236.

Ison, D. C., & Szathmary, K. J. (2016). Assessing academic integrity using SafeAssign plagiarism detection software. *The Collegiate Aviation Review International*, *34*(1). http://dx.doi.org/10.22488/okstate.18.100466

Kidwell, L. A., & Kent, J. (2008). Integrity at a distance: A study of academic misconduct among university students on and off campus. *Accounting Education: An International Journal*, *17*(S1), S3–S16.

Lancaster, T., & Clarke, R. (2016). Contract cheating: The outsourcing of assessed student work. In T. Bretag (Ed.) *Handbook of Academic Integrity* (pp. 639–654). Springer. doi:10.1007/978-981-287-079-7_17-1

Martin, D., Rao, A., & Sloan, L. (2011). Ethnicity, acculturation, and plagiarism: A criterion study of unethical academic conduct. *Human Organization*, *70*(1), 88–96.

Martin, D. E., Rao, A., & Sloan, L. R. (2009). Plagiarism, integrity, and workplace deviance: A criterion study. *Ethics & Behavior*, *19*(1), 36–50.

Malesky, Jr., L. A., Baley, J., & Crow, R. (2016). Academic dishonesty: Assessing the threat of cheating companies to online education. *College Teaching*, *64*(4), 178–183

Medway, D., Roper, S., & Gillooly, L. (2018). Contract cheating in UK higher education: A covert investigation of essay mills. *British Educational Research Journal*, *44*(3), 393–418.

Moxley, J. M., & Eubanks, D. (2016). On keeping score: Instructors' vs. students' rubric ratings of 46,689 essays. *WPA: Writing Program Administration-Journal of the Council of Writing Program Administrators*, *39*(2), 53–80.

Peterson, J. (2019). An analysis of academic dishonesty in online classes. *Mid-Western Educational Researcher*, *31*(1), 24–36.

Ramzan, M., Munir, M.A., Siddique, N., & Asif, M. (2012). Awareness about plagiarism amongst university students in Pakistan. *Higher Education*, *64*(1), 73–84.

Rogerson, A. M., & Basanta, G. (2016). Peer-to-peer file sharing and academic integrity in the Internet age. In T. Bretag (Ed.) *Handbook of academic integrity* (pp. 273–285). Springer. doi:10.1007/978-981-287-079-7_17-1

Scanlon, P. M. (2003). Student online plagiarism: How do we respond? *College Teaching*, *51*(4), 161–165.

Scanlon, P. M., & Neumann, D. R. (2002). Internet plagiarism among college students. *Journal of College Student Development*, *43*(3), 374–385.

Shahid, U., Farooqi, S., Ahmad, R., Shafiq, Z., Srinivasan, P., & Zaffar, F. (2017). Accurate detection of automatically spun content via stylometric analysis. In *2017 IEEE International Conference on Data Mining*. pp. 425–434. IEEE.

Stuber-McEwen, D., Wiseley, P., & Hoggatt, S. (2009). Point, click, and cheat: Frequency and type of academic dishonesty in the virtual classroom. *Online Journal of Distance Learning Administration*, *12*(3).

Sutherland-Smith, W. (2016). Authorship, ownership, and plagiarism in the digital age. *Handbook of Academic Integrity*, 575–589.

Walker, J. (2010). Measuring plagiarism: Researching what students do, not what they say they do. *Studies in Higher Education*, *35*(1), 41–59.

Watson, G. R., & Sottile, J. (2010). Cheating in the digital age: Do students cheat more in online courses? *Online Journal of Distance Learning Administration*, *13*(1).

Wilkinson, J. (2009). Staff and student perceptions of plagiarism and cheating. *International Journal of Teaching and Learning in Higher Education*, *20*(2), 98–105.