Plagiarism by Adult Learners Online: A case study in detection and remediation*

Christine Jocoy - California State University, Long Beach, USA
David DiBiase - The Pennsylvania State University, USA

Abstract
Detecting and combating plagiarism from Web-based sources is a concern for administrators and instructors involved in online distance education. In this paper, we quantify copy-and-paste plagiarism among adult learners in an online geography course offered through Penn State’s World Campus Geographic Information Systems (GIS) certificate program. We also evaluate the effectiveness of an “expectation management” strategy intended to discourage adult learners from unintentional violations. We found that while manual methods detected plagiarism in only about 3 percent of assignments, Turnitin.com revealed a 13 percent plagiarism rate among the same assignments. Our attempts to increase awareness and manage expectations decreased infractions measurably, but not significantly. In contrast, Turnitin.com substantially improved our ability to detect infractions. We conclude that raising awareness and managing expectations about plagiarism may be worthwhile, but is no substitute for systematic detection and vigilant enforcement, even among adult learners.

Keywords: Plagiarism; academic integrity; cheating; online; e-learning; adult education

Introduction
Educators have always been concerned with upholding standards of academic integrity among individuals engaged in scholarly pursuit. For many institutions of higher learning, academic integrity is viewed “as a basic guiding principle for all academic activity” (Penn State University Faculty Senate, 2000). Members of intellectual communities, such as universities, are expected to value honesty, trustworthiness, and civility and to behave accordingly (McCabe and Pavela, n.d.; Princeton University, 2003; York University, 2005). These standards of behavior are meant to ensure “that work done is one’s own and that the work of others is properly recognized” (College of Agricultural Sciences, 2005). As a basic tenet of scholarly activity, educators have a responsibility to foster and maintain standards of academic integrity, which requires engaging students in the development of moral reasoning (Kohlberg and Hersh, 1977). One approach to effective maintenance of standards and the advancement of students’ moral development is the detection and remediation of specific violations.
Plagiarism is one type of violation of academic integrity. The Council of Writing Program Administrators states that “Plagiarism occurs when a writer deliberately uses someone else’s language, ideas, or other original (not common-knowledge) materials without acknowledging its source” (as cited in Quinn, 2006). With the proliferation of digital source material on the Web, plagiarism has received renewed attention, particularly among administrators and instructors involved in online distance education (Groark, Oblinger, and Choa, 2001; Heberling, 2002; Hickman, 1998). Some observers believe that the Internet makes it easier for students to plagiarize (Harris, 2004; Saulnier, 2005). Underwood and Szabo (2003) find that students with more exposure to Internet use in assignment preparation self-reported greater willingness to engage in copy-and-paste plagiarism (i.e., copying word-for-word without citing the source). Hinman (1999) goes so far as to suggest that we soon will witness an increase in academic dishonesty as universities offer more courses through online distance education.

As instructors of an online distance education course, we share these concerns and chose to examine the extent of Internet plagiarism in five offerings of our online course between July 2003 and June 2004. Our course requires students to be active Internet users, including creating an online portfolio in which they post their assignments as webpages. This paper presents the results of our investigation of plagiarism prevalence, detection, and remediation among adult learners in an online course. We focus specifically on copy-and-paste plagiarism, the copying of another author’s language word-for-word without proper citation.

Defining and Identifying Plagiarism

The Council of Writing Program Administrators’ definition of plagiarism reveals several contingencies which complicate the enforcement of academic integrity in higher education, particularly in regards to adult education. Intentionality is one contingency. Writers’ uses of the works of others are not always deliberate. Infractions may result from mismatches between the ethical norms of the academy and the workplace (Martin, 1994), or simply from hasty and incomplete adaptation of passages copied and pasted from digital sources for reference purposes. Some might consider the latter an example of poor writing rather than plagiarism because it did not involve intentional cheating. Nonetheless, it is difficult to ascertain intentionality through detection except in extreme cases. (Extreme cases would include copying entire or large portions of papers written by someone else or papers purchased from term paper mills.) Lack of proof of intentionality may reduce the penalties for offenders, but detecting writing that gives the impression of being plagiarized remains important for offering instructive remedies. Unintentional violations of the rules do not mean that plagiarism has not occurred. Regardless of the degree to which an infraction is intentional or substantial, plagiarism violates an original author’s intellectual property rights. The “fair use” provision of US copyright law does allow quotations and paraphrasing of original works without authors’ permission. When original authorship is not properly acknowledged, however, such uses constitute copyright infringement.

A second contingency is the degree of culpability. Researchers have identified different forms of plagiarism (Cabe, n.d.; Martin, 1994). Copying another writer’s language (e.g., directly quoting word-for-word several sentences of common-knowledge materials) or poor paraphrasing (e.g., substituting a few synonyms for original text without significantly changing it) may be judged less substantive infractions than an attempt to pass off someone's
else’s idea as one’s own. Some educators suggest that concern with plagiarism should be more about teaching students to appreciate the development of knowledge, acknowledge intellectual contributions of other scholars, and represent the process of building on existing knowledge in academic writing and less about violating rules and copyright laws (Howard, 2003; Hunt, 2003; Martin, 1994).

Because academic integrity involves the development of behavior that reflects moral values, educators’ responsibility for addressing plagiarism may go beyond shielding students from copyright infringement. Students whose use of other authors’ works is constrained only by the perceived risk of detection and punishment may fail to recognize the relevance of the rights of others (Kohlberg and Hersh, 1977). By seeking more than mere compliance with what may seem to students to be arbitrary rules, therefore, educators enforce academic integrity in order to advance students moral development (Dark and Winstead, 2005). These considerations, which complicate the identification of plagiarism and enforcement of standards, affect instructors’ ability to quantify, detect, and prevent incidences of plagiarism. The following sections review previous research on these three issues.

### Quantifying Plagiarism Prevalence

Previous studies report widely varying percentages of cheating prevalence (Crown and Spiller, 1998; Ercegovac and Richardson, 2004; Lathrop and Foss, 2000; Whitley, 1998). Crown and Spiller (1998) attribute this to variation in the measurement of cheating along three dimensions: actual observance versus self-reporting of instances, type of cheating behavior evaluated, and time period in which cheating occurred. First, most research quantifies cheating by means of self-reporting in student surveys (CAI research, 2005; Ercegovac and Richardson, 2004; Whitley, 1998). Actual observations of cheating produce different results, usually much lower estimates of prevalence (Crown and Spiller, 1998; Karlins, Michaels, and Podlogar, 1988). Second, Whitley’s 1998 review of the literature on college cheating indicates that cheating behavior is most frequently defined as cheating on examinations followed by estimates of total cheating, cheating on homework, and plagiarism.

Estimates of total cheating typically include a variety of types of cheating behavior (e.g., copying from another student’s exam or quiz; using notes during an exam; turning in a paper written by someone else; falsifying citations; failing to cite source material; unauthorized collaboration on homework (McCabe and Trevino, 1996). The wide range of cheating rates reported in the literature (i.e., from 9 to 95 percent of students for total cheating and from 3 to 98 percent for plagiarism as reported in Whitley (1998) suggests that the type of cheating behavior explains some of the variation in incidence levels (Crown and Spiller, 1998). Third, the bounding time frame for incidence occurrence, either for observance or survey self-reporting of cheating behavior, affects prevalence rates. Incidence rates for cheating on homework assignments over the course of a semester will differ from rates of plagiarism on one assignment or self-reporting of any incident of cheating during a college career (Crown and Spiller, 1998).

Less common in the literature are studies that examine the prevalence of plagiarism separately from other forms of cheating (Whitley, 1998). Most assessments of the prevalence of plagiarism alone among students rely on self-reporting in surveys (CAI research, 2005; Ercegovac and Richardson, 2004; Scanlon and Newman, 2002). Such surveys are often
conducted within individual universities (Ercegovac and Richardson, 2004), but a few studies sample from multiple universities (CAI research, 2005; Scanlon and Newman, 2002). A few studies measure actual rates of plagiarism detection among student assignments as a way to gauge prevalence (Braumoeller and Gaines, 2001; Karlins, Michaels, and Podlogar, 1988; Soto, Anand, and McGee, 2004). Given the effort involved in systematic cheating detection, studies reporting observed rates assess fewer plagiarism types, shorter time frames, and smaller samples of students or assignments at a single university. This may reflect the finding that instructors are reluctant to report student cheating and therefore have no desire to set up systematic procedures for detecting it (CAI Research, 2005; Ercegovac and Richardson, 2004). Differences in the types of plagiarism detected may account for some of the variation in plagiarism rates obtained in the three studies.

Detecting Plagiarism

Instructors’ ability to detect plagiarism has improved dramatically in the last decade. Prior to widespread use of the Internet, detection was limited to the manual efforts of instructors. For example, in the Karlins, Michaels, and Podlogar (1988) study, two people manually compared papers submitted during the current and preceding semester that contained citations of the same sources for verbatim copying or poor paraphrasing. Today, increased Internet use makes it both easier for students to copy-and-paste from online materials and for instructors to detect infractions (Braumoeller and Gaines, 2001; Tenbusch, 2002). Free online search engines such as Google allow instructors to track down copied phrases, while commercially available plagiarism detection software and online services (e.g., EVE; Turnitin.com) compare individual student papers to Web documents and/or to essay databases to find and report instances of matching text.


We compared our ability to detect Internet plagiarism within our students’ assignments using two different methods. This allowed us to evaluate and compare the effectiveness of automated and manual methods of plagiarism detection.

Preventing Plagiarism

While recognizing the importance of plagiarism detection, we are also interested in prevention. Our own experience with student infractions supports the conclusion of Center for Academic Integrity (CAI) researchers that many students have yet to develop a clear sense of appropriate Internet use in written assignments (CAI research, 2005). Other researchers found actual observed infractions to be associated with a lack of knowledge about plagiarism (Soto, Anand, and McGee, 2004). Many educators view explicit plagiarism instruction as the best means of prevention (Conradson and Hernandez-Ramos, 2004; Harris,
 Recent case study research provides support for the effectiveness of incorporating plagiarism instruction into individual courses. Soto, Anand, and McGee (2004) found that students who received no explicit plagiarism instruction plagiarized twice as often as those who participated in active instructional activities such as class discussions of definitions of plagiarism, review of Turnitin.com plagiarism reports, and exercises requiring students to identify instances of plagiarism in example essays. We also tested the impact on infraction rates of providing explicit plagiarism instruction in the form of an expectation management strategy introduced prior to students’ preparation of assignments.

Case Study

We analyzed 429 assignments prepared by students enrolled in five sections of Geography 482: The Nature of Geographic Information between July 2003 and June 2004 for signs of plagiarism. Geography 482 is a required first course in both the Post-baccalaureate Certificate and Master of Geographic Information Systems (GIS) degree programs offered by Penn State’s Department of Geography and delivered through the University’s World Campus. The course has been offered four times a year since 1999, attracting 40 to 100 students per offering. The purpose of the course is to introduce students to the field of GIS and to orient them to online learning. Students meet the latter objective by creating and maintaining a webpage portfolio (e-portfolio) of their course assignments.

Enrollees tend to study part-time while maintaining full-time employment. Ages of students enrolled during the study period ranged from 25 to 87; the median age was 41 years. Students were located throughout the US and to a lesser extent the world. Sixty-six percent were male. Most had undergraduate degrees obtained through previous higher education. Many were practicing GIS professionals seeking formal education in geography and GIS in support of career advancement. Others pursue the certificate in hopes of entering the field.

In our course, copying and pasting text from Internet websites was the form of plagiarism that students were most likely to engage in, given the requirements for the assignments. We selected two of three project assignments to examine for signs of plagiarism. (...) We were concerned with copy-and-paste plagiarism in this introductory course because assignments did not ask students to make evaluative or critical arguments, nor were they required to make original research contributions.

Methods

This study investigated three aspects of plagiarism. The first objective was to quantify actual rates of copy-and-paste plagiarism in student assignments. The second objective was to compare our ability to detect plagiarism manually with automated methods provided by plagiarism detection software. Manual methods were used during the process of grading the assignments during each of the course offerings. After the completion of the courses, the assignments were re-evaluated for plagiarism using an automated detection service. The third objective was to contrast assignments prepared by students who were given minimal plagiarism instruction to assignments completed by those receiving explicit instruction. This before-and-after comparison revealed the extent to which explicit instruction reduced occurrences of plagiarism.
Quantifying plagiarism using manual detection

In instructions for the course assignments, plagiarism was defined generally as the unacknowledged use of ideas, words, or illustrations produced by other authors. Students were given a link to the definition of plagiarism used by Penn State’s College of Earth and Mineral Sciences (College of EMS, 2002).

During the process of grading, each assignment was evaluated for upholding principles of academic integrity using manual methods for plagiarism detection. The grader looked for common signs of copy-and-paste plagiarized work: inconsistent citation styles, lack of citations in long passages, awkward formatting, use of dated language, use of difficult vocabulary and terminology, and irregularities of diction and style (Harris, 2004). Suspect text was checked against work cited in bibliographies and through Google searches for copied phrases and sentences. We employed a strict standard for classifying text as plagiarized. An assignment was considered to contain plagiarism if it included: 1) at least one sentence copied verbatim from an online source without the inclusion of quotation marks and a citation or; 2) two or more poorly paraphrased sentences that also lack citations. In the second case, poor paraphrasing was identified as sentences including too many of the author’s actual words or phrases and/ or the author’s original sentence structure.

Quantifying Plagiarism Using Automated Detection

We used the same criteria described above for classifying text as plagiarized for both manual and automated methods of detection. We are reasonably confident that manual and automated methods detected copy-and-paste plagiarism in a similar fashion because all of the papers identified as containing plagiarism using manual methods also were detected using Turnitin.com. The automated reports generated by Turnitin.com calculate a percentage of copied text, but we did not use these measures in our determination of plagiarism. Instead, we looked at each report and the text from original sources in side-by-side comparisons. This is because Turnitin.com cannot distinguish automatically between plagiarized text and properly cited direct quotations. Instructors must interpret the results documented in the reports, which can still be quite time-consuming. Nonetheless, Turnitin.com speeds up the process of finding copied text and finds it through more systematic searching than can be undertaken using manual methods.

Results

The use of an automated plagiarism detection service noticeably improved our ability to find and document instances of copy-and-paste plagiarism. Manual detection missed nearly 4 in 5 cases of plagiarism. Manual methods enabled detection of papers containing high levels of plagiarism, such as a paragraph copied word-for-word. Detection using Turnitin.com was more exact, uncovering instances where students copied just one sentence or several long phrases word-for-word. Compared with manual methods, the Turnitin.com search engine proved to be more systematic in searching the Web and precise in matching assignment text to its original source.

While Turnitin.com increased our plagiarism detection abilities appreciably, the expectation management strategy only marginally reduced rates of plagiarism. We observed a 3.5 percent...
decrease after its implementation, but this improvement was not statistically significant. We did not find any difference in the number of pre-quiz and post-quiz repeat offenders (students who plagiarized on both assignments) using Turnitin.com for detection. We did find, however, that of the five students caught (using manual detection methods) and penalized for using plagiarized material on the first assignment, none repeated the violation on the third assignment.

**Discussion**

Our finding of a 13 percent rate of plagiarism is in line with rates obtained from other studies that measured actual infractions. Given that previous studies quantified plagiarism among traditional undergraduate students, we also conclude that plagiarism rates among adult learners may not be lower than those for younger students. Additionally, our finding of a relatively low rate of plagiarism supports the notion that self-reported rates from survey questionnaires are likely to be higher than those obtained through actual detection of plagiarism (Crown and Spiller, 1998; Karlins, Michaels, and Podlogar, 1988). Instructors and researchers should heed the warnings of Crown and Spiller (1998) and make sure to account for the important bounding conditions of prevalence studies in their interpretation of rates.

The use of Turnitin.com improved our ability to detect cut-and-paste plagiarism measurably. While the automated process of checking papers was not necessarily faster than manual checking, it was certainly more thorough, enabling us to adhere and enforce to a stricter definition of plagiarism. (...) While our findings lead us to conclude that plagiarism search engines are effective assessment tools, some questions remain about their robustness. Braumoeller and Gaines (2001) specifically tested the detection accuracy of EVE software with a test paper known to contain plagiarism by running the paper through the system three times. Nonetheless, we believe the automated methods to be superior to manual ones, at least for copy-and-paste forms of plagiarism.

We did not see a significant reduction in plagiarism with the use of our expectation management strategy. Nonetheless, there was a small improvement in accordance with previous research that documents significant improvement (Soto, Anand, and McGee, 2004). We view expectation management as generally good practice, especially considering evidence that associates a lack of knowledge about plagiarism with higher rates of incidence (Soto, Anand, and McGee, 2004) and with student anxiety about committing offenses unintentionally (Ashworth, Bannister, and Thorne, 1997). By using an academic integrity quiz to assess student understanding, we go beyond basic written (or verbal) instruction, which by itself produces marginal, if any, deterrence to plagiarizing (Braumoeller and Gaines, 2001) and is not likely to provide the kind of instruction that furthers students’ moral development.

We concede that our expectation management strategy does not provide students with a complete understanding of the dimensions of plagiarism nor a full appreciation of the role of proper citation in the development of knowledge and intellectual pursuit. Nonetheless, we do believe that the expectation management strategy combined with detection and enforcement using Turnitin.com emphasizes to students the importance of academic integrity and reinforces the values of institutions of higher education.
Conclusion

Educators are keenly interested in ways to assure the academic integrity of their students’ work. This fact was highlighted for us while working with colleagues at Leeds University and the University of Southampton on a project concerned with developing e-learning resources for geographic education. (…) Like Braumoeller and Gaines (2001), however, we conclude that expectation management alone is no substitute for rigorous enforcement of academic integrity standards. Based on the research reported here and in previous studies, we are convinced that even the most vigilant grader is likely to overlook many, if not most, infractions. (…) We expect that effective detection and enforcement will lead to a higher level of compliance with academic integrity standards in this introductory class, as well as in the classes that follow in our certificate and master’s programs.

References


*This paper has been abridged for use in the Pre-sessional Programme - Course D for educational purposes only. Students do not have permission to reproduce this material in any form.

This work is licensed under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/by/4.0/).