This article reports the results of a trial of automated detection of term-paper plagiarism in a large, introductory undergraduate class. The trial was premised on the observation that college students exploit information technology extensively to cheat on papers and assignments, but for the most part university faculty have employed few technological techniques to detect cheating. Topics covered include the decision to adopt electronic means for screening student papers, strategic concerns regarding deterrence versus detection of cheating, the technology employed to detect plagiarism, student outcomes, and the results of a survey of student attitudes about the experience. The article advances the thesis that easily-adopted techniques not only close a sophistication gap associated with computerized cheating, but can place faculty in a stronger position than they have ever enjoyed historically with regard to the deterrence and detection of some classes of plagiarism.
THE ISSUE: TECHNOLOGICAL ASYMMETRY

University professors have been losing a technological arms race with their students over plagiarism for many years. Since the end of the days when term-papers and other assignments were hand-written, students willing to misrepresent others’ work as their own have exploited an expanding array of tools. From the simple cutting-and-pasting from uncited sources or each other’s work, to the use of keyword-searchable collections of term papers available for purchase, students have at their disposal a broad palette of quite powerful techniques.

That these methods have become commonplace for students and notorious among their instructors has not led many faculty to employ commensurately sophisticated technologies for detection. A good deal of concern is now directed at the plagiarism crisis in universities: honor codes are being adopted or expanded, reports are being commissioned, and journalists are producing a stream of news stories. But in most cases, the techniques instructors actually use for identifying plagiarism and thereby deterring it remain frankly old-fashioned, unchanged for the most part from the good old days of pen-and-paper assignments. Even in cases where classes feature course web sites, electronic presentations, and other applications of contemporary technology, instructors typically fall behind their own students in degree of technological sophistication when it comes to matters of cheating. This gap in sophistication between students and their instructors is one of many pressing issues created by the rapid evolution of information technology in the university.

This article describes and discusses a classroom trial of pattern-matching software capable of detecting certain categories of plagiarism. The article describes the origins of the trial in dissatisfaction with traditional techniques, the approach to using cheating software, the results, considerations of the dual aims of deterring and detecting cheating, and future steps that will follow.

The thesis advanced here is that readily available technologies are capable of doing more than simply closing the sophistication gap between students and instructors. We show that use of simple technologies can in some respects place faculty in a stronger position than they ever enjoyed historically, even in the days when cumbersome pen-and-paper assignments using printed rather than electronic sources meant cheating was more difficult and labor-intensive for students.
One of the conceits of technological advance has long been that the solution to any human costs of innovation is simply more innovation. But it is not often that social or moral problems created by use of technology can be remedied well by the use of additional technology. Plagiarism may be one exception. The reason is straightforward. The fundamental technological problem we see is a simple asymmetry in the use of existing information technology. Virtually all students build, manipulate, and can share electronic versions of their course work, but they then submit to their instructors only paper copies. Those paper copies then narrowly circumscribe instructors’ power to identify the process that went into their creation. So long as this asymmetry exists, students will have the upper hand in cheating, regardless of whatever exhortations and lessons in honesty are directed at them by their universities. The simple transition to electronic submission of student work rather than the transfer of paper is the foundation for remedying this imbalance as well as for other pedagogic advances.

THE FAILURE OF TRADITIONAL APPROACHES

It now appears clear that the tepid response of most universities to students’ cheating opportunities constitutes a serious problem. The nationwide conversation about what appears to be an epidemic in academic dishonesty continues unabated (Zernike, 2002). A study by the Center for Academic Integrity (Groark, Oblinger, & Choa, 2001) found that almost 75% of college students own up to some form of academic dishonesty. And the level of cheating is on the rise: the same study found that in 1963, 11% of students at medium to large universities said they collaborated on assignments when it was expressly not permitted, and this figure had increased to 49% by 1993. McCabe and Treviño (2002) reported that a number of institutions have adopted honor codes for the first time or revised existing documents in response to the perceived crisis in cheating. However, the problem continues in the face of traditional responses, and the Internet in particular has added a level of complexity to the problem to which universities are struggling to develop effective responses.

Of course many categories of cheating exist, and the demands they create on instructors vary. We are concerned here with cheating on term papers in large survey classes. In particular, we are concerned with the sharing of work between students, whatever its origins. Large introductory classes are paradigmatic of some of the cheating problems for a number of reasons.
For one, assignments are often graded by more than one person, making it difficult to check for similarity among papers. Students can readily surmise that no one person will be able to read hundreds of papers and so be in a position to identify duplicate work. Also, because so many students take these large classes every year, and because assignments are often at a fairly basic level, the potential supply of plagiarizable sources is great. Paragraphs or even pages of material may be clipped with ease from web sites. Worse, complete papers may be available from classmates, fellow athletes, roommates, the Greek system, and the wide array of Internet-based term-paper vendors.

In the past, faculty offering such courses at our institution have employed a number of approaches in varying combinations to thwart some categories of academic dishonesty. These include:

- Highly specific, directed, term paper questions intended to reduce the probability that students are able to find or purchase complete papers meeting the requirements of the assignment;
- Rotation of term paper assignments every year or two to confound the collection of usable term paper files by fraternities or other student groups;
- Assignment of different term paper questions to each Teaching Assistant’s (TA) students, so that opportunities for cross-TA sharing of papers are reduced;
- Selective testing of specific phrases from suspicious-looking papers using Internet search engines to look for possible sources of material.

These techniques are at the very best modest in their power to detect and deter cheating. Our conclusion from many years in this course is the same as that of many faculty in other classes: these offer little promise in the face of the powerful tools available to students. In some cases these techniques are difficult to apply and can even compromise pedagogic quality, such as in the case of supplying narrowly crafted questions or varying assignments across Teaching Assistants or instructors.

For several years, some faculty, including one of the present authors, have employed a rudimentary electronic technique for identifying gross cases of duplicate material in papers across Teaching Assistants. We have required
each Teaching Assistant to enter into a spreadsheet several phrases from
every student’s paper. These might, for example, be the last five words in
the first and last paragraphs, or the first words in the second sentence of
particular paragraphs, or some other varying algorithm. With these spread-
sheets in hand, comparisons can then be made across students using simple
spreadsheet functions. Papers flagged by this technique as sharing phrases
are then read by hand. This technique is enormously labor-intensive, and is
likely only to detect unsophisticated, large-scale copying between two or
more papers. It is easily defeated by students who are diligent in making
slight modifications between plagiarized papers, and our experience in this
class is that students submitting shared work literally always make slight
changes throughout in order to disguise the duplication.

Despite its severe limitations, this technique has consistently identified an
average of one case of plagiarism each time it is employed in a course of
three to four hundred students. The fact that such a loose net catches some
instances of cheating has been unsettling, since it implies that many other
instances of cheating are likely going undetected.

ELECTRONIC SUBMISSION OF PAPERS AS THE FOUNDATION:
SETTING UP THE TRIAL

Our experiences with these various techniques, especially with nominal
electronic comparison of tiny samples from student papers using spread-
sheet functions, made clear how desirable it would be to conduct a test of
a technique involving the full text of every student’s paper. In that case,
where student work is available electronically to the instructor, a world of
opportunities is open. In theory, the electronic text of a student paper can
be compared with other students’ work from the class in question, with
historical archives of past papers, with papers of other instructors, including
those at other institutions, and in fact with any electronically searchable file,
including all those on the open Internet.

In recent years a number of products have become available to assist in this
type of paper comparison. News reports of their success are circulating
(Stoerger, 2002) and commercial vendors have recently become assertive
in touting the capabilities of their products (Kane, 2002). Some scholarly
reports of their efficacy have also been reported (e.g., Braumoeller &
Gaines, 2001). One approach is the pattern-matching or comparison engine that runs on a local computer at the school or university. WORDCheck is a commercial example costing at the time of this writing between $100 and $1,000, depending on the size of the installation. Also, some faculty have written their own utilities for comparing student work, since the basic function of comparing text between files is so simple technically. The best-known case, reported by Schemo (2001), occurred at the University of Virginia, when in 2001 physics professor Lou Bloomfield found 158 papers with matching phrases from an introductory class. His pattern-matching software eventually resulted in the dismissal of 45 students and the revocation of three degrees. Another approach to comparing papers involves commercial services to whom instructors may submit student papers. Some of the more popular examples are Turnitin.com, Eve2, Copycatch, and Paperbin.

The goal of our examination was to adopt and test a technology for term-paper comparison that would meet several criteria: it must be inexpensive, it must be readily expandable and adoptable by other faculty, and it must provide a highly level of flexibility and control to the instructor. We selected Copyfind 2.1, the freeware utility developed by Lou Bloomfield of the University of Virginia. This software is available free, runs on a local computer host rather than through a commercial service or other off-campus third-party, and is adaptable and highly flexible.

We chose for our trial a large introductory political science course on American government taught by one of the authors. It is typical of introductory surveys, enrolling from 350 to 600, depending on the term. It attracts a wide range of students who take it as a general education course, a gateway course into the political science major, or as a prerequisite for a state teaching credential. The course usually requires that students submit one term paper of about ten pages or, in some years, two analytic essays of five pages each. It is also directly comparable to the course involved in the report on plagiarism-detection software by Braumoeller and Gaines (2001).

The course has for several years, employed web-based syllabi and writing assignments, so adding the capacity for electronic submission of papers was straightforward. We prepared a simple file-uploading web page connected to the course web site. Here students ready to submit their papers entered their name, student identification number, e-mail address, TA’s name, and the file name of their paper. Completing the transaction was
similar to “checking-out” at a retail web site. Students ready to transmit their papers received a presubmission verification notice of the information they had entered and the name of the file they were about to submit; they then clicked on a “Submit” button which transferred the file to our host computer, and finally they received from us a postsubmission confirmation notice that we had received their paper. We received 590 term papers averaging a little under 10 double-spaced pages in length.

ANALYSIS AND RESULTS

The analytic method employed in the Copyfind software is pair-wise comparison of phrases at the sub-sentence length. Each paper is parsed into a list of phrases of a few words in length, and then each paper’s phrase list is compared one-by-one against every other paper’s list. The output is a count of the number of identical phrases for each possible pair of student papers, regardless of the order or location of the matching material. No matching phrases indicates the absence of plagiarism in a pair of papers, while very large numbers of matching phrases indicates probable plagiarism, which can then be confirmed by a manual reading of the papers in question.

As an example of this technique, consider a paper containing the following sentences: “Instructed delegate representation is an approach to democracy in which elected officials follow the expressed wishes of their constituents. In contrast to trustee representation, they set aside their own expertise, information, and judgments...”

Parsed into every possible sequential three-word phrase, this text produces a list of phrases beginning as follows:

Instructed delegate representation

delegate representation is

representation is an

is an approach

an approach to
Instructed delegate representation is an approach to democracy in which elected officials follow the expressed wishes of and so forth (etc.)

Selection of an appropriate phrase length for comparison is a simple optimization problem. Set the phrase length too short and a large number of nonmeaningful matches between papers will be identified; set it too long and otherwise identical phrases in which students have altered just one word will be rejected as different. We approached the problem empirically by making multiple runs, varying the phrase length each time, until we found a useful value. 4

Our set of 590 papers produces 173,755 possible paper-pairs. Despite the large number of phrases in each paper and the large number of pairs, running the set required only about five minutes of time on a dual 800 MHz PowerMac G4 running MacOS X 10.2.1.

We found that using any but the longest phrase length, a large number of papers shared some substantial number of phrases with every other paper, since many students use stock phrases or idioms, arrive at a few similar sentences coincidentally, begin paragraphs in similar ways, or quote the same source material here and there. We found that several dozen innocent
matching phrases was typical. Many paper pairs with up to two or three hundred matching phrases also turned out to be innocent of plagiarism upon reading by hand.

We found that paper pairs where the number of matches exceeded 500—in one case the figure was over 2000—turned out to be plagiarized. Against the background of some dozens and even a few hundred matching phrases, papers sharing more than a paragraph or so stand out with a large spike in the number of matching phrases. Figure 1 displays each unique number of paper pair matches that appeared in the comparisons, from 0 matches to 2214 matches.\(^5\)

\[\text{Figure 1. Distribution of identical phrase matches among paper pairs}\]

In our 590 papers, we found one pair and one triple with vastly greater than the baseline number of random matching phrases. These represent the four paper-pairs at the extreme right of Figure 1, with more than 500 identical phrases. We read the top two-dozen or so paper-pairs by hand, representing matches greater in number than 300, and found just these four to be plagiarized. In the pair, roommates’ papers shared an identical section amounting to about 20% of the paper, appearing at the end. This represented a moderately severe but unequivocal case of plagiarism. Before the campus Student Conduct Committee, one student confessed in this case to taking material from his roommate’s paper without the roommate’s knowledge. The plagiarizing student was given an F in the course by the instructor and was suspended from the university for two quarters.
In the other case, three papers shared massive verbatim text, running to about 90% in one pair. This was the paper flagged by the software with 2214 identical phrases. Two of the students involved admitted using the work of the third student while it was in the possession of a fourth student, who had been given the original paper to proofread and who served as an intermediary in sharing the work. In this case, three students were suspended from the university for two quarters: both plagiarizing students and the intermediary, who was convicted of aiding and abetting their plagiarism.

**STRATEGY OF DISCLOSURE: DETERRENCE OR PUNISHMENT?**

Having the capacity to compare student papers thoroughly raised an important question about the purpose of detecting plagiarism. In general, the apprehension of rule-breakers serves two social purposes: it punishes transgressors and it deters some people from proscribed behavior. The way that rule-enforcement activities are publicized may affect the balance that is struck between punishment and deterrence. If members of a society are completely unaware that detection and punishment of rule-breaking are occurring, little or no deterrent effect is accomplished. At the same time, where deterrence is perfect, then no punishment will occur. We deliberated at some length about how much to reveal to our students about the nature of the efforts we would make in detecting cheating.

We surmised that a full disclosure to students would likely have a very powerful deterrent effect. Much of the literature on plagiarism written by educators and other scholars, as noted by Stoerger (2002), endorses a strategy of deterrence in all cases, and recommends full disclosure. In our case, we anticipated that the deterrence offered by disclosure of our software would likely extend to students who might be habitual or committed cheaters in other classes, but who would recognize that their odds of detection in this particular class would be extremely high—perhaps 100%. We therefore expected that if we disclosed fully we would, ironically, be likely to catch no one in the act of cheating. We also anticipated a different effect, namely that the more information we revealed about our techniques the more crafty a few students might be in attempting to disguise efforts at plagiarism. In particular, knowing or guessing the phrase length used in our comparison would open a back-door to evading the software, though at the cost of considerable effort. At the same time, we realized that not revealing our
technique would forfeit some of the deterrence, especially for students who might be causal, infrequent, or noncommitted cheaters.

Our goal was to deter students directly but also to detect and punish any cases of the most negligent or incorrigible efforts at cheating. We therefore adopted a middle strategy between full disclosure and complete concealment. We wanted to deter students who might be tempted to cheat but also detect and punish students who might be committed to cheating. We call this soft-deterrence, and it worked as follows.

We implemented this strategy as follows. We gave repeated oral and written warnings and exhortations to students not to cheat, indicating that we would be making efforts to catch them and that punishment would be severe. These notices appeared in the syllabus and in the paper assignment itself, and were made in class on the first day of the term and again the day before the paper was due. We made a special point to say that the instructor would be “on the lookout for students submitting papers that are essentially the same or that share sections.”

We also provided a special web page for students defining plagiarism and providing resources supplied by our university and several others about academic honesty and mechanics of quoting and citing properly. All this took place in the context of instructions that papers would be submitted electronically. We stopped short, however, of telling students that we would be comparing their papers electronically.

Our intention was that this thorough set of warnings would serve to deter casual cheaters but not committed or foolhardy ones. We likened this to a public announcement that police would be out in force to detect speeders and drunk drivers over a holiday weekend, but without an indication of just where police officers would be located.

Our outcome of two cases of plagiarism strikes us as an endorsement of the soft-deterrence strategy. We can say with assurance that of our 590 students only these five were involved in sharing text with one another. We can also say with assurance that no two students used any common source such as a fraternity or sorority file copy of a paper or a commercial paper service, since these would have appeared in our analysis as shared papers (though the source would be unknown to us). This strikes us impressionistically as a somewhat lower overall level of cheating than we would have expected in a
class this large without the announcement and warnings and whatever inherent deterrent effect the electronic submission itself might have had.

Our test of this technique also included a student survey. We included three measures in the standard course-evaluation survey given to students on the last day of class, several days after they had submitted their term papers. The first measure inquired about how students liked electronic submission of their papers compared with the traditional technique of turning in a printed copy. Students were given a seven-point feeling thermometer in which the middle score indicated they had no preference between electronic and paper submission, one end of the scale indicated a strong preference for submitting electronically, and the other end a strong preference for submitting the traditional way. This permitted students to express a preference for either and indicate how strong that preference is, or to express no preference.

The results show that about three quarters of students either prefer electronic submission or are indifferent, with the remaining quarter preferring traditional physical submission of papers. Specifically, 57% of students favored electronic submission, 15% reported no preference, and 28% favored traditional submission. On the seven-point scale, the modal response was a score of 7, which represented the strongest level of preference for electronic submission. As this was the first trial of this technique in this course, and for a large majority of students was probably the first experience of submitting any college paper this way, we interpret these results as a quite strong level of student acceptance.

Our other two measures asked whether students felt any deterrent to cheating from the instructor’s exhortations in class and from the use of the electronic submission technique. We interpret the results with a grain of salt for two reasons. First, the findings of any survey inquiring about illicit activity are to be read with caution. Second, the survey asked students to consider a counterfactual question about their own behavior, namely whether they might have been more inclined to cheat under other course circumstances. Such hypothetical questions are far less reliable in surveys that are straightforward factual questions.

The first of these deterrence questions posed the following. “Students are often tempted to borrow parts or all of a paper from another student. Did the warnings that were given in this course that the instructor would be on
the lookout for shared papers have any effect on you?” A seven-point scale permitted answers ranging from “no effect” to “kept me from borrowing.” (Note that we did not differentiate among “no effect” answers between situations where a student cheated despite the warning and where a student would not have cheated regardless of the warning.) The results of this question are striking: 66% of students reported that the warning had at least some effect on them. Over a quarter of students—27%—chose the strongest category of response indicating that the warning prevented them from borrowing work from another student. Figure 2 shows the tri-modal response.

![Figure 2. Self-reported effect of instructor warning on propensity to cheat](image)

The second deterrence question addressed the electronic submission procedure. It asked: “What about the fact that you submitted the paper electronically—did this have any effect on whether you decided to borrow parts or all of a paper from another student?” Response distribution here shows the same overall shape, but with substantially more students reporting “no effect” and fewer—just 15%—saying that this technique kept them from borrowing. Figure 3 shows the responses.
In this case as in Figure 2, the question elicited some indecision, which is reflected in the central mode at the mid-point of the scale. We interpret this and the other responses near it as an indication of students being unclear about the situation and wishing to register uncertainty about whether they were affected by the arrangements in the class. This strikes us as unsurprising. We conclude that something between a large plurality and a small majority are simply determined (not to cheat or to cheat), a nontrivial minority of up to a quarter are likely tempted to cheat but can be deterred, and the rest—a substantial number—are up for grabs morally.

The satisfying lesson we draw from these data are that the moral authority of exhortation against cheating is apparently at least as strong and probably stronger than the implicit threat of detection from electronic submission. We conclude that these data are supportive of our strategy: it appears that the approach we used in this course had some deterrent effect along with its proven capacity actually to detect cheating.
DISCUSSION OF RESULTS AND NEXT STEPS

We enjoyed two unexpected benefits from this trial of plagiarism-detection technology. First, having a password-protected, central repository of all student papers proved useful in several ways aside from cheating detection. The professor had ready access to student work and could browse papers at will. Also, TAs with questions during grading could correspond with the professor who had immediate online access to each paper. The other benefit was the time-stamping and logging of papers. In a class this large, questions traditionally arise about when a paper was submitted, just how late a late paper was, and even whether a student submitted or not. Our electronic submission technique produced an electronic log with time-stamps for each paper, and this proved useful in several such cases when students came forward with problems. In one case, for instance, a student claimed to have submitted his paper late by just an hour or so, which would have fallen within our grace period; the log showed the paper a half-day late and immediately settled the matter.

The largest problem we encountered in this trial involved printing out physical copies of the papers for use in grading. On the principle of innovating just one step at a time, we chose not to require the ten TAs in this class to grade papers electronically and return marked-up copies to students by way of e-mail or some other means. This meant we needed hard-copies of the papers. We also chose not to require students to submit both electronic and physical copies, because of the problems to which that might be susceptible: a student submitting versions that differed, a student submitting one on time but one late, a student submitting one but not the other, and so forth. Instead, we sorted, printed, and stapled the electronically submitted papers centrally using a high volume printer and then distributed these to each TA for grading.

Central printing produced several problems. We required students to submit in MS Word or WordPerfect format, but about 30 (5%) failed to do so and these papers required special attention. We also experienced loss of certain kinds of formatting, especially “smart quotes” in some papers, apparently due to version differences in word processors.

About five students (<1%) encountered technical problems submitting their papers. We operated a live telephone hot-line for problems in the three hours before the paper was due, and handled a couple of problems after the
due date without penalty to the students. These proved in all cases to be operator error at the student end: students attempting to submit twice, or in one case attempting to submit to us an entire directory of files from her computer.

We consider this trial a success and will continue to use Copyfind. We intend to extend the system within this large course in three ways. First, we will retain student papers over time and will compare future student submissions not only with one another but with the archive of past papers. Because the assignments in this class rotate every two or three years, the archive will never grow unwieldy in size.

Second, we aim to employ a standard Internet search engine such as Google to test for matches with our phrase lists from student papers. This will require some additional coding to submit the search strings and then read back and tabulate search results for each paper. While several uncertainties exist about the time needed for such searches and for the interpretation of results, the principles involved are straightforward. Open searching on the Internet will allow detection of plagiarism from a wide variety of sources, including in some cases even commercial term paper mills, because these often publicly post short excerpts from their stock of papers.

Third, we will move eventually to electronic grading of papers to eliminate the need for printing altogether. Having electronic copies of papers makes this possible, and the utilities for marking up text with highlighting, marginalia, and even audio notations are now widely available as part of most standard word processors. In this case, we will accept student papers electronically, copy each one, run one copy through our plagiarism detection routines while grading and marking up the other. We will then return the graded papers to students by way of e-mail.

We will also extend our experimentation within the campus by arranging trials in more courses within more departments. These trials, if successful at detecting and preventing plagiarism, will also help identify the nature and extent of services necessary to support the routine use of the system in whatever courses faculty members desire.

Because the core engine is freeware, the system certainly is usable at no direct cost to other institutions. The authors are glad to share their experience and suggestions with other potential users.
IMPLICATIONS

The implications of this simple trial of electronic detection of plagiarism are sobering: if all student work was tomorrow checked electronically for plagiarism, schools and universities would likely be overwhelmed with academic dishonesty cases. Our own university now handles a few dozen cases each year through its central student ethics office, in the absence of widespread use of electronic detection methods by faculty. Extrapolating from our trial in one class to the hundreds of courses offered here each year suggests that in the short run the number of cases of dishonesty caught and prosecuted could easily grow by an order of magnitude were electronic techniques widely used by faculty. The immediate consequences in labor costs and student punishments and suspensions or dismissals from widespread use of electronic detection of plagiarism would be frankly staggering.

The longer-term consequences of general use of electronic detection methods, however, might be substantially more desirable: a growing expectation by students that cheating is wrong and that it is more often detected and punished than not. The ease with which we were definitively able to detect cheating and probably to deter it in our trial does not merely indict students for reckless behavior. It suggests that university faculty bear some considerable responsibility for the current crisis in academic dishonesty for not taking more positive steps to keep pace with students’ own use of technology.

We interpret the crisis in academic dishonesty as more than simply a behavioral phenomenon in which growing numbers of students are committing acts of misrepresentation and dishonesty for reasons that are imponderable or that are external to what teachers and administrators do. As a number of observers including Donald McCabe, founder of the Center for Academic Integrity at Duke University, have pointed out a central problem is the cavalier attitude toward cheating that seems to have grown among students (Groark, Oblinger, & Choa, 2001; Slobogin, 2002). Why have attitudes toward cheating changed in recent decades, as so many observers have noted?

One answer has to do with the ease with which students can cheat undetected using new technology and the languid response by instructors at catching them. Where rules are trivially easy to break and the makers of rules extend themselves little to detect violations, a strong if implicit signal is sent that the rules are not important. To the extent that universities have failed to
respond adequately to cheating by failing to maintain a deterrent through
detection or to apply honor codes, they bear some substantial responsibility
for the crisis.

References

Braumoeller, B.F., & Gaines, B.J. (2001). Actions do speak louder than words:
Deterring plagiarism with the use of plagiarism-detection software. PS On-
line. Retrieved June 27, 2003 from http://www.apsanet.org/PS/dec01/brau-
moeller.cfm
mills, anti-plagiarism tools, and academic integrity. EDUCAUSE Review.
erm0153.pdf
Academe/02JF/02jftoc.htm
Schemo, D.J. (2001). University of Virginia hit by scandal over cheating. New
news/04/05/highschool.cheating/index.html
http://www.web-miner.com/plagiarism
Zernike, K. (2002). With student cheating on the rise, more colleges are turning
to honor codes. New York Times (November 2). Retreived June 27, 2003,
from http://www.nytimes.com

Notes

1. A small sample of these paper mills includes Cheathouse.com (for-
com, LazyStudents.com, MightyStudents.com, SchoolSucks.com, and
Termpapersrus.com.

2. Questions have been raised about violation of students’ privacy and in-
tellectual property rights in the cases of commercial paper comparison
services. Although a number of educators and schools continue to find this objectionable, the prevailing view at this point seems to be that this practice is ethically acceptable and we agree within certain constraints. As for intellectual property concerns, the primary purpose of course work is not commercial gain but fulfillment of course requirements, so expectations of intellectual property rights are not high among undergraduates. Nonetheless, whatever intellectual property value might inhere in a written work is in no way diminished by a technique of private comparison with the work of others. Indeed, it might be argued that a procedure that validates student work as original could enhance intellectual property value. Privacy concerns are related. High standards for privacy are necessary for both school officials and their contractors.

3. See http://plagiarism.phys.virginia.edu/Wsoftware.html

4. We do not report the phrase length here publicly in order to avoid creating an invitation to students to attempt efforts at defeating the software.

5. Note that many possible match levels appeared frequently: for instance, 33,000 paper-pairs had 10 or fewer matching phrases, about 5000 paper-pairs had 45-55 matching phrases, and just 4 paper-pairs had 500 or more matches. These 500 were cases of plagiarism.