Article: "Actions Do Speak Louder Than Words: Deterring Plagiarism with the Use of Plagiarism Detection Software"
Author: Bear F. Braumoeller, Brian J. Gaines
Issue: December 2001
Journal: ***PS: Political Science & Politics***

Actions Do Speak Louder than Words: Deterring Plagiarism with the Use of Plagiarism-Detection Software

Bear F. Braumoeller; Brian J. Gaines

PS: Political Science and Politics, Vol. 34, No. 4 (Dec., 2001), 835-839.

Stable URL:
http://links.jstor.org/sici?sici=1049-0965%28200112%2934%3A4%3C835%3AADSLTW%3E2.0.CO%3B2-J

PS: Political Science and Politics is currently published by American Political Science Association.

Actions *Do* Speak Louder than Words: Deterring Plagiarism with the Use of Plagiarism-Detection Software

Bear F. Braumoeller, *Harvard University*
Brian J. Gaines, *University of Illinois at Urbana-Champaign*

Introduction

In recent years, the availability of online source material and online papers has increased instructors' concerns regarding plagiarism in the classroom. Many instructors do not realize, however, that the digital revolution has also created a niche for fast and (at least somewhat) reliable plagiarism-detection software.

In the spring semester of 2000, we were slated to teach two 90-student sections of Political Science 100: Introduction to Political Science at the University of Illinois at Urbana-Champaign. We recognized the potential for a controlled experiment to assess whether or not (a) explicit warnings not to plagiarize and (b) the overt use of plagiarism-detection software had any impact on rates of plagiarism. As ancillary benefits, the experiment also allowed us to estimate the efficacy of such software and the extent to which plagiarism, whether casual or blatant, constitutes a problem in such classes.[1] The detailed results follow, but in brief, the answers to these questions are:

- Warning students not to plagiarize, even in the strongest terms, appears not to have had any effect whatsoever. Revealing the use of plagiarism-detection software to the students prior to completion of an assignment, on the other hand, proved to be a remarkably strong (though still not absolutely perfect) deterrent.
- In our first trial, about one out of eight papers was deemed problematic due to either casual or blatant plagiarism. Of these, nearly all fell into the casual category. While we cannot with confidence establish an upper bound on percentage of papers demonstrating plagiarism, one-eighth serves as a fairly solid lower bound.
- Existing plagiarism-detection software is not perfect, but its success rate is high enough to merit use in a wide range of classroom situations.

Bear F. Braumoeller *is an assistant professor of government at Harvard University and an affiliate of both the Weatherhead Center for International Affairs and the Davis Center for Russian Studies. Prior to July 2000, he was an assistant professor at the University of Illinois. His primary research interests include international security, political methodology, and the study of Russian foreign policy.*

Brian J. Gaines *is an associate professor of political science at the University of Illinois at Urbana-Champaign. His teaching and research interests include electoral systems and voting behavior, political methodology, and the political histories of the Anglo-Atlantic democracies.*

Background

Concern over cheating by college students is, of course, not new. For decades, survey results on the incidence of various kinds of academic dishonesty have provided eyebrow-raising statistics. From the 1960s through the 1990s, it has not been difficult to find surveys in which over half of student respondents acknowledge firsthand experience with cheating, broadly defined (Maramark et al. 1993).

High-tech detection, on the other hand, is comparatively new. Plagiarism-detection algorithms became something of a hot topic within the specialized world of computer programming in the 1980s (e.g., Berghel and Sallach 1985), but over a decade passed before entrepreneurs spotted the commercial potential of plagiarism-detection software for the wider educational market.[2] In the past five years, education journals have been full of blurbs and "puff" pieces about Plagiarism.org, Integriguard.com, EVE and WordCHECK. A few lengthier works have revisited the broader question of how professors can discourage plagiarism at the outset. Vernon et al. (2001) propose various common-sense remedies for plagiarism—e.g., make the penalties clear, let students know that you know about online paper mills, provide targeted non-generic instructions for papers, and so on—and briefly describe detection methods, alerting readers to the existence of a few specialized sevices.

They do not, however, report any experience with the accuracy, thoroughness, or utility of any of these services. Indeed, there seem to be few detailed reports of firsthand experience with these services. Hereafter, we detail our experience with one piece of software and draw a few conclusions about its utility and its broader implications for the shifting balance of technology in the classroom.

Procedure

We gave one essay assignment to both sections of POLS 100. The students were asked "to write about five pages (1,500–2,000 words) relating any major theme or topic in the first nine chapters of the course textbook to a recent event in world politics." The assignment was intentionally broad; our purpose was not to encourage plagiarism, but rather to remove impediments to it in order to assess student behavior when topical constraints are few. A written warning about plagiarism appeared on section D's (Professor Gaines's) assignment sheet, but not on section C's (Professor Braumoeller's). In all other ways the assignment sheets were identical. Professor Gaines also issued a stern, verbal warning against plagiarism when handing out the assignment; Professor Braumoeller did not.

We searched online sources such as Yahoo! for plagiarism-detection software and services, trying to find one that would not run afoul of the University's legal stipulations—a nontrivial requirement.[3] We finally arrived at a satisfactory solution. We created course websites at Blackboard.com (<www.blackboard.com>), and students handed in their papers to the TAs electronically via Blackboard.com's "Digital Dropbox" feature. We then ran the papers through a program called the Essay Verification Engine, or EVE (<www.canexus.com/eve>), version 2.1. Trial versions are available, and individual licenses cost $19.99. The program's author describes its functions as follows:

"EVE fragments the essay based on a number of rules and uses these fragments to conduct searches in a variety of areas. . . . The essay stays on your computer, where your EVE software performs matches and statistical analysis against material it retrieves from the web" (Personal communication, Professor Braumoeller, 28 March 2000).

Because initial trial runs demonstrated variability in the program's performance, we checked each paper three times. All papers that were flagged in any run were then examined by our teaching assistants, to determine whether the papers actually included problematic content. Using Internet and library resources, we followed up on any leads that EVE provided. We revealed the results of the experiment to the class prior to the due date for their second required essay, and we repeated the procedure on that essay to determine whether the knowledge that such software was in use would affect rates of plagiarism.

Results

1. In deterrence, actions speak louder than words.

We designed the first part of the experiment to gauge the impact of verbal and written warnings on rates of plagiarism. The results can be described only as depressing. In POLS 100C, 10 of 78 papers (12.8%) were ultimately deemed problematic; in POLS 100D, 9 papers out of 73 (12.3%) were caught. Although there is a very slight difference, it is highly likely ($c^2 = 0.0083$, $Pr = 0.927$) that these two distributions were drawn from the same population, implying that written and verbal warnings had a negligible effect, at best.

That the first papers had to be graded and handed back provided us with an opportunity for a second experiment. Since students would eventually discover that some papers had been marked down, we decided to reveal the experiment to the students during lecture. We could then run their second assignments through the same software to ascertain whether public revelation was a deterrent.

Students' reactions were difficult to gauge[4] and probably mixed: no one likes feeling tricked, but those students whose grades had improved had little reason to complain. Even students who received deductions for poor citation practice or outright copying were in no position to revolt: the penalties they incurred were quite light in terms of the overall course grade, and we made a point of erring on the side of leniency when classifying citations as adequate. Professor Braumoeller took the opportunity to demonstrate how honest students' grades were pushed down, relative to the class mean, by plagiarism. To bring the point home in concrete terms, he illustrated the grading curve before and after penalties had been assessed, and demonstrated that an honest student who had initially received a B on the paper would receive a B+ once the grading curve had been adjusted. In fact, the net result of penalizing problematic papers was an average increase of one third of a grade for all papers not deemed problematic. As this discussion drew to a close, quite a few students were slowly nodding their heads. Most seemed torn between anger (at having been the subject of such an experiment) and happiness (at probably having received a free, if minor, grade increase). Interestingly, no one complained about not having been warned, despite these students having been in the "no warning" group. They did, however, have quite a few questions about the precise definition of plagiarism. The overwhelming majority seems to have gone to great pains in the next paper to stay well within accepted boundaries.

By chance, we subsequently received aid from the student newspaper, the *Daily Illini*, which called the political science department (on a slow news day, perhaps) and asked whether any of the professors would like to talk about the topic of plagiarism. Professor Braumoeller revealed the details of the experiment, and the article appeared on the front page not long thereafter (April 18, 2000). Any POLS 100 students who had missed lecture and had not read the online lecture notes, therefore, could read about the experiment in the student newspaper.

The results of this deterrent were much more encouraging. On the second assignment, only one student submitted a paper that was quite clearly problematic.[5] The student in question later testified that she had taken a job that required her to work during the class lecture period and that she had managed to miss all of the other warnings, including the newspaper article and the online lecture notes, which contained the details of the experiment.[6] We therefore conclude that the deterrent effects of actually checking for plagiarism are quite impressive, though one should never assume that every single student has gotten the word—no matter how widely it is disseminated.

2. At this stage, plagiarism-detection software is useful.

We conducted initial trial runs using EVE on a mock-up "essay" consisting of unrelated paragraphs taken from 15 online essays, and the results were grounds for cautious optimism. Although estimates of plagiarism rates varied (see below), EVE usually flagged most passages from most websites. Running student papers through EVE highlighted an unanticipated bonus: EVE flagged many online news sources, magazines, etc., and comparatively few student essays that have found their way into the public domain.

Though we obviously would have been happier with a consistent 100% detection rate on the trial essay, an inability to flag every source on every run does not constitute a grave impediment. Even a low probability of detection can create a credible deterrent.

3. The results of plagiarism tests should not be taken to be definitive.

The percentage of the test document flagged by EVE in the first four trial runs varied to an uncomfortable degree (28.64%, 5.6%, 4%, 37.88%).[7] Because the middle two trials were the only ones to report plagiarism rates under 25% and were conducted in very close temporal proximity, we conducted three more trials on the same essay over a longer time period and found more encouraging results:

Trial 1. 3/27/00 (3:55 pm) 44.86% plagiarized 16 websites
Trial 2. 3/28/00 (11:56 am) 42.48% plagiarized 18 websites
Trial 3. 3/28/00 (1:20 pm) 52.67% plagiarized 21 websites

Part of EVE's function is to report the percentage of the document that can be attributed to other sources. Plagiarism-detection software faces a very difficult challenge, especially when trying (as EVE does) to catch paraphrases and slight changes in wording. Therefore, a fair number of false positives should be expected. Such software also cannot distinguish between passages that have been stolen and those that have been properly cited—almost half of the papers flagged by EVE demonstrated proper citation practice.

There are both advantages and disadvantages in the inclusion of a report feature that flags suspicious text by looking for passages that approximate the text in question. The main advantage, of course, is that the time-honored practice of tinkering with included text to avoid detection is less safe. One of our students, for example, took a section of text—describing the passage of a treaty—nearly verbatim from a source document without attribution, making minor changes here and there: 90 signatories became 100, for example. EVE caught

the passage. On the other hand, the program's decision rules cannot cover every possible scenario, and, thus, flagged passages *must* be investigated thoroughly. Our sample document contained nothing but a name, a title, and plagiarized material. In one run, EVE correctly flagged the first paragraph as plagiarized but then curiously extended the flagged area backward to include the title and the student's name.

Some papers originally flagged as only minimally plagiarized turned out, upon closer inspection, to contain virtually no original work. In one case, a student compiled a paper from four printed sources, one of which was also available online. Once EVE had flagged passages from the latter source, the structure of the rest of the paper strongly suggested that we needed to investigate the remaining sources. We discovered that nearly the whole paper was copied verbatim from the sources cited in the bibliography, with only a few trivial word substitutions.[8]

These results suggest that multiple plagiarism tests across different time periods are warranted and that any "percentage plagiarized" feature may be misleading. The software should *not* be used to estimate such percentages but rather to flag papers for further inspection.

4. Plagiarism is a problem, though quite possibly not as pervasive as most people think.

Because it was possible for students to plagiarize and evade detection by EVE, we cannot present a precise estimate of the incidence of plagiarism. We can, however, establish a lower bound: 12.6% of our students (to be more precise, 12.583%) "represent[ed] the words or ideas of another as [their] own," which fits the University's official definition of plagiarism.[9] Actual plagiarism rates under these conditions are almost certainly higher. It would be difficult to argue that they are lower.

For comparison, we surveyed students about how much plagiarism they believe is typical on paper assignments in a large introductory lecture class. We asked about what percentage of student papers in "classes such as the one you are in now" fall into each of these three (mutually exclusive and exhaustive) categories:

Original. The work constitutes the student's original effort, and any ideas or arguments not original to the student are properly cited.

Casual plagiarism. Passages within the paper bear too close a resemblance to passages in the source material to constitute original work, and even though the original source is cited somewhere in the paper, the method of citation that the student uses does not make clear that the passage in question is not original.

Blatant plagiarism. Part or all of the paper is taken from another source without attribution.

We surveyed students after they had submitted their papers, but before we had revealed to them that we were screening the papers with plagiarism-detection software.[10] Figure 1 shows the distribution of student responses. The x-axis represents the estimated percentage of papers that are wholly original, the y-axis is the percentage of papers containing casual plagiarism, and, since the three categories are exhaustive, the distance from the hypotenuse along a 45-degree angle represents the expected percentage of blatant plagiarism. (Since the diagram is slightly unconventional, we have labeled a few cases for clarity.) The two highest estimates for blatant plagiarism were 40%, while the maximum estimate for casual plagiarism was a whopping 80%. Across both sections, mean values were: 60% original; 32% casual; 8% blatant.

Although having some plagiarism in approximately one paper out of every eight is unacceptable (especially as a lower bound), these surveys suggest that students may be even more pessimistic (or cynical) than is warranted. Of course, we hesitate to describe our results from the screening process as the "truth" about the amount of plagiarism. The software is plainly not perfect, we had no way to detect how many papers came from fraternity files and other such offline sources, and we made any no effort to determine if any students were recycling papers by submitting identical works in multiple classes without permission. Nonetheless, we remain struck by the fairly low levels of detected plagiarism, as against student estimates. Even if we assume that we caught only 50% of the actual instances, the adjusted amount would then still be below the mean estimated amounts of plagiarism.

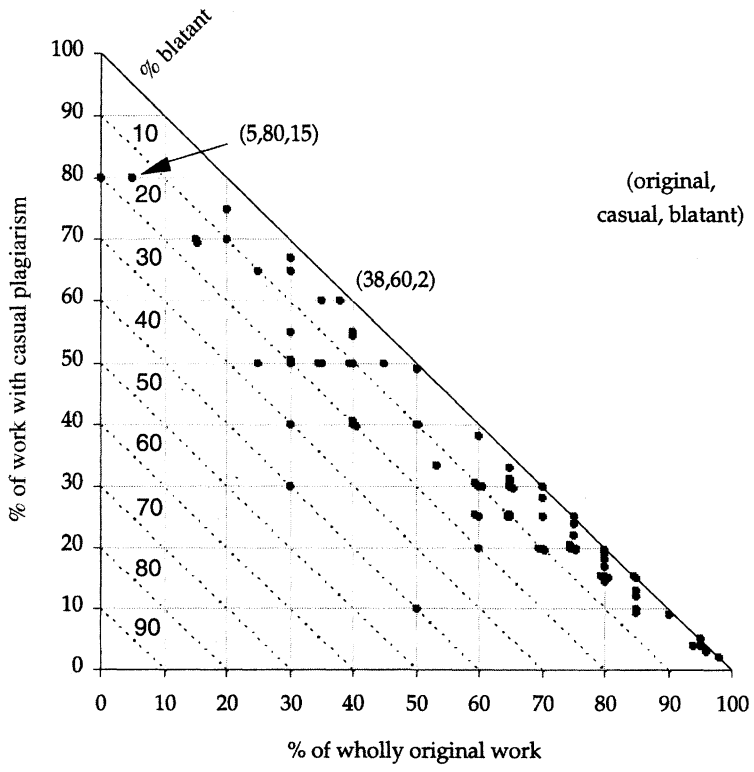
5. Blatant plagiarism is not as common as "casual" plagiarism.

Using the same scheme as the survey item (described above), we classified papers as (wholly) original; containing instances of casual plagiarism; or demonstrating blatant plagiarism.

About 20% of the papers failed to receive a completely clean bill of health from EVE. However, because in many cases all passages highlighted as copied were in fact quotations properly cited, we quickly eliminated several cases and reduced the number of problematic papers to 19 (out of 151). The vast major-

Figure 1

Students' Estimates of Plagiarism Rates



ity of these cases of plagiarism fell into the second category. Such cases are annoyingly ambiguous. Optimists will assert that they result from bad citation practice and the failure of high schools to train students to produce original written work. Pessimists will claim that they represent the work of strategic plagiarists who plan to fall back on ignorance as an excuse if caught. Students, inevitably, will profess ignorance of convention. In four instances, students had submitted papers that turned out to contain so many passages directly or almost directly copied from published sources without proper citation that we classified them "blatant."

6. Plagiarism substantially harms honest students' grades.

Our interest in this project stems from our desire to ensure that our students can take our classes secure in the knowledge that they will not be penalized for honesty. This experiment demonstrated that the great majority of them would have been disadvantaged had we not engaged in monitoring activities. On the first assignment, papers containing significant amounts of uncited or unreferenced material were initially graded as if they were unproblematic, but then discounted to reflect the percentage of unoriginal material. A paper that received an 88 but was found to be about 50% unoriginal, for instance, received a grade of 44. The results illustrate the extent to which plagiarism harms honest students' grades, as mentioned in section 1.

7. Ironically, paper mills may, in the long run, make plagiarism more difficult.

Our test document consisted of passages taken from 15 essays available online at various websites.[11] As mentioned, EVE proved capable of detecting material taken from those sources. In POLS 100C, for example, EVE initially flagged 18 student papers and produced references to a total of 85 websites that are remarkably diverse in origin. Some were traditional news sources (washingtonpost.com), some were online versions of traditional printed sources (books, magazine articles, treaties), some were online dictionaries or encyclopedias, and so on. Only one paper mill was included. Moreover, that site was flagged only because the online paper and the student's paper both quoted sections from the U.S. Constitution. As the quotations were properly attributed to the original source, this single paper-mill reference turned out to be a red herring.

Our suspicion is that paper mills in general are not a grave problem. Our inspection of a random sample suggests that the papers they contain are of middling quality at best; students may reach the same conclusion. Moreover, the existence of student paper mills has created a niche for plagiarism-detection software. It is increasingly common for printed sources (newspapers, magazines, and books) to appear on the Internet, both on parent websites and in student essays. Even papers from fraternity files can be scanned and sold to online paper mills by students wishing to make fast money. The surprising result is that plagiarism-detection programs are increasingly capable of catching passages taken from printed, rather than online, sources. For example, EVE found that one of our students had plagiarized from an article in *The New Republic*, which is not archived online but is often quoted directly in online articles and essays. It remains to be seen whether selective and strategic alteration of words is an effective strategy for evading plagiarism detectors. But the practice of buying and selling essays may in the long run make plagiarism *more* difficult: *any* passage or paper might turn up somewhere on the Internet, even if the original source was printed rather than electronic.

Conclusions and Recommendations

This experiment was relatively limited in scope: it involved only one of many plagiarism-detection systems and took place in one class at one university. Nevertheless, its findings should be of considerable interest to instructors. Although simply warning students about plagiarism seems to have no discernable deterrent effect, advertising the use of plagiarism-detection software seems to concentrate minds wonderfully. Overall, we found that while a few students engage in intentional academic dishonesty, far more were unclear on the rules against plagiarism, but, paradoxically, had received enough lectures on it that they simply "tuned out" any warnings. The challenge for the educator is to deter the first group and to motivate the second to pay closer attention. Plagiarism-detection software seemed to serve both functions quite well.[12]

We stress one caveat: the majority of the cases that we encountered, although they met the university's definition of plagiarism, fell into a gray zone somewhere between proper citation practice and outright theft. Software is likely to unearth quite a few such cases, so anyone contemplating its use would be well advised to prepare by, for example, providing handouts with examples of proper and improper citation practice, making students aware of relevant university regulations, and so on. Prior to the experiment, we thought it wise to discuss the nature of the penalties to be assessed with our deans, who obliged us by quite clearly explaining what penalties they deemed appropriate. We strongly recommend this course of action to anyone contemplating the use of plagiarism-detection software.

Inevitably, some students will develop countermeasures to the kind of software that we have discussed herein. The simplest method would be to take material exclusively from printed sources, but as our experiment demonstrates, even printed sources are not perfectly safe from detection. The practice of quoting and citing those sources could evolve as a more foolproof way of avoiding punishment. In fact, if plagiarism-detection efforts become more common, students who set out to plagiarize might of necessity find themselves engaging in actual research.

Notes

1. We are grateful to the University of Illinois for providing funding to cover both the expense of the software and overtime hours for our two hardworking teaching assistants, Oana Armeanu and Jennifer Romine; we are also most grateful to Armeanu and Romine for their time and effort. From *PS*, three anonymous referees, Sheilah Mann and Blake Brunner contributed very useful suggestions.

2. Interestingly, or perhaps obviously, it is not only students who succumb to the temptation to misuse the Internet. The new ease of accessing other sources by Internet seems also to have created a rash of plagiarism-related firings and an attendant debate on what to do in the circles of professional journalism (see, e.g., Robertson 2001). The newspaper or magazine editor and the professor do not, of course, face identical problems. Detection of cheating in the classroom allows fairer grading and assists in teaching proper citation practice and/or alerting miscreants to the dangers of cutting corners. Newspapers, far more concerned with their own reputations for accuracy than with assigning proper credit (or lack thereof) to contributors, do not appear to be focusing on software to screen submissions so much as better enforcement of existing fact-checking procedures.

3. The details are complex, but the major sticking points had to do with issues of intellectual property. Plagiarism.org, for example, reserves the right to retain student essays so that future essays can be compared to them, but students' essays are their own property; the combination of these two facts raised red flags in the University's legal department, and we were never able to arrive at an acceptable solution.

4. Normally we might have considered using

course evaluations to assess student reactions to the experiment, but in this case the measure they provided was too coarse to be of much use—we had never taught the class before and so had no baseline for reference, the papers in question were only a small part of the overall course, etc. For what it's worth, we did not find our evaluations on the whole to be appreciably more positive or negative than what we would otherwise have expected. The night before Section D's final exam, Professor Gaines's office was burglarized, but we hesitate to connect that incident to this experiment.

5. One reviewer has raised the question of whether carrying through with the threat of monitoring by lowering students' grades was necessary in order to achieve the desired deterrent effect. We wish we had thought of that question ourselves during the experiment, but we didn't. In fact, the potential penalties for plagiarism on the second assignment were quite dire—the Deans had indicated approval of any sanction that we deemed appropriate, including dismissal from the University. We do not know whether lesser sanctions, or even simple monitoring, would have been as efficacious.

6. The student was subsequently asked to leave the University, due in large part to her academic performance—though her role in this experiment did not go unnoticed. She has since petitioned for reentry, fulfilled the deans' requirements, and reenrolled.

7. We are grateful to Rick Braumoeller for his assistance with these plausibility probes. We contacted Matt Hunter, the creator of EVE, to ask about the variability of these results. Without having witnessed them, he could only speculate that one or more of the servers that EVE uses to do its sleuthing might have been down—a hypothesis consistent with both the timing of the low-percentage runs (very close together) and the consistency of the other results. Because the program relies on other sites, which are usually but not always available, multiple detection runs separated by at least an hour seem prudent. We settled on three because initial runs suggested that virtually all of the material that was going to be flagged would be caught in the first three runs; more would in all likelihood have been a waste of time.

8. Here we are grateful for the impressive sleuthing skills of Delinda Swanson, who used Internet search engines to locate an article that was particularly difficult to find, and to Lynne Rudasill, who used library software to do the same—and to inform us that it could be found in the next building.

9. We would be remiss, and ironically so, if we did not cite the University of Illinois's Code of Policies and Regulations Applying to All Students, Rule 33, Section I, subsection D.

10. Not atypically for a large lecture that satisfies distribution requirements, POLS 100 generated fairly low attendance rates. Response rates also not being perfect, we received 82 surveys.

11. The websites were <www.bignerds.com>, <www.cyberessays.com>, <www.essayglobe.com>, <ww.dlc.fi>, <www.cheater.com>, <www.essaydepot.com>, <www.planetpapers.com>, <www.coshe.com>, <www.members.xoom.com>, <www.essays.simplenet.com>, <www.oppapers.com>, <www.papercamp.com>, <www.netessays.net>, <www.chuckiii.com>, and <www.geocities.com>.

12. Ideally, the rewards of research would render obsolete the need to engage in this sort of activity, but we are skeptical about that possibility. We permitted these students to write on virtually any topic, and encouraged them to pick a subject that they found stimulating. Even given this degree of latitude, a substantial percentage of the students turned in work that, going by the letter of the law, could have resulted in suspension or expulsion.

References

Berghel, H.L., and David L. Sallach. 1985. "Computer Program Plagiarism Detection: The Limits of the Halstead Metric." *Journal of Educational Computing Research* 1:295–315.

Maramark, Sheilah, and Mindi Barth Maline. 1993. *Academic Dishonesty Among College Students*. Washington, DC: Office of Educational Research and Improvement.

Robertson, Lori. 2001. "Ethically Challenged." *American Journalism Review* 23 (2): 20–29.

Vernon, Robert F., Shirley Bigna, and Marshall L. Smith. 2001. "Plagiarism and the Web." *Journal of Social Work Education* 37 (1): 193–196.