

Textbook plagiarism in PSY101 General Psychology: incidence and prevention

J. Stephen Mansfield, D.Phil.
Department of Psychology, SUNY Plattsburgh

steve.mansfield@plattsburgh.edu
<http://faculty.plattsburgh.edu/steve.mansfield>

1.0 Introduction

This study investigates plagiarism by General Psychology students. Informal observations indicate that students often use phrases from the class textbook in their written homework assignments. This type of plagiarism may not be intentional (e.g. students could inadvertently copy from the text if they are lazy or sloppy during their note taking so that they forget whether their notes are direct quotes, or paraphrases in their own words). Intentional or not, it is important that plagiarism is detected and inhibited. If a paper containing plagiarized material receives a good grade, then the author receives tacit approval of his or her writing strategy. Plagiarism behavior that is reinforced is more likely to be repeated.

I have used a computer-based plagiarism analysis to scan homework assignments to find cases where students have copied directly from the class textbook. This system is a deterrent to would-be intentional plagiarizers, but more importantly, it also serves as a teaching tool to illustrate to students how frequently they copy phrases from the textbook when they should have used their own words.

4837 assignments were analyzed over two semesters. I have used these analyses to address the following questions: (a) What is the incidence of plagiarism? (b) What factors predict plagiarism? (c) Can a computerized plagiarism detection system reduce plagiarism?

2.0 Methods

2.1 Participants

Writing assignments were analyzed from a total of 176 students from 3 sections of General Psychology (2 sections in Fall 2001, and one section in Fall 2002). Students completed their homework assignments as part of their regular course work.

2.2 Homework assignments

Students answered questions based on the day's chapter from the class textbook (Shaffer & Merrens, 2001). The questions asked about methodological details and findings from the research described in the chapter. Homework was submitted via email prior to each class. The analyses in this paper are based on 4,836 assignments ($M=24.9$ per student, $SD=6.7$).

Spelling errors can mask potential matches between a student's work and the text. To avoid this, students were instructed that they would be penalized for poor spelling. Nevertheless, spelling errors were still present in many assignments. This will act to artificially reduce the incidence of plagiarism reported in this study.

2.3 Analysis

The students work was analyzed with a computer program to find strings longer than 25 characters that were identical to the textbook. The matching algorithm is based on Greedy-String Tiling (Wise, 1996). Prior to analysis, the student's work was normalized by removing non alphabetic characters and separating each word by a single space.

2.4 Feedback

The amount of plagiarism feedback given to the class varied in the different sections due to differences in teaching methodology and differences in the availability of technology. One section received no formal feedback about the plagiarism analysis (this section provides a baseline level for the incidence of plagiarism), one section was warned during the 8th week of the semester that plagiarism was being checked automatically. From then on, students who had excessive amounts of copied text in an assignment were graded zero for that assignment and privately warned to not do it again. The third section (which was taught a year later) received plagiarism feedback privately via the web for every assignment. Portions of work that were identical to the text were highlighted. Students who had excessive amounts of copied text in an assignment were graded zero for that assignment and privately warned to not do it again.

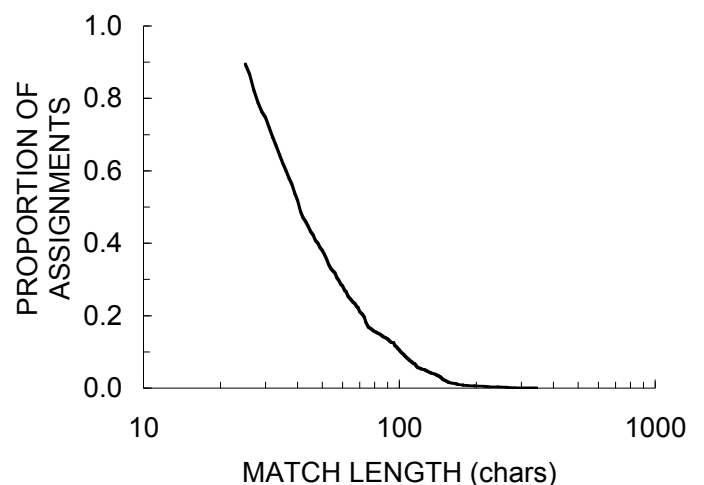


FIGURE 1. Proportion of homeworks submitted with a matching text string longer than the specified match length (e.g. 10% of assignments contained at least one matching string that was 100 characters or longer).

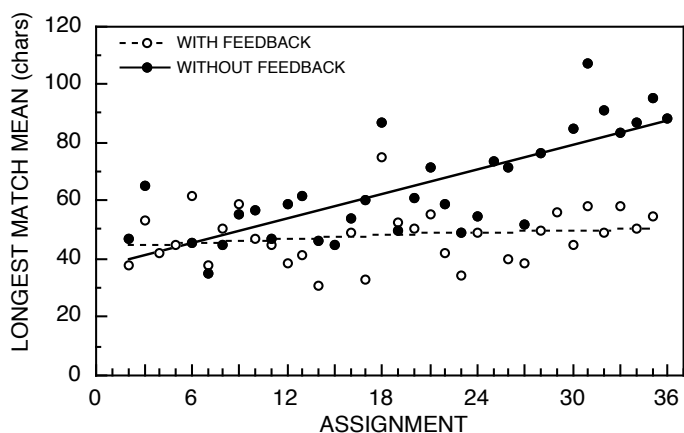


FIGURE 2. Changes in plagiarism during the semester. The longest matching string from all students homeworks was averaged across the class for each assignment. This is plotted as a function of the assignment. Assignment 18 occurred at mid term. Data are shown for a section that received no feedback, and a section that received feedback from mid term onwards.

3.0 Results & Discussion

3.1 Overall incidence of plagiarism.

Figure 1 shows the proportion of student assignments that contained plagiarized text as a function of the length of the copied string. These data are from the class that received no feedback about plagiarism. Figure 1 allows us to experiment with different plagiarism criteria. Arguably, a matching sequence of 2 or more words (approximately 10 characters) could be considered plagiarism. However, such a stringent threshold seems unreasonable because the text often included multi-worded scientific terminology (e.g. *Minnesota multiphasic personality inventory*). Adopting a criterion match-length of 25 chars (approximately 5 words) classifies 90% of all assignments as plagiarized! Even with a longer match criterion, the number of plagiarized assignments is remarkably high: a 50-character (10 words approx.) criterion classifies 38% of the assignments as plagiarized, and a 100-character (20 words approx.) criterion classifies 10% of the assignments as plagiarized. These data indicate that a large number of assignments routinely exceed a very liberal plagiarism criterion.

3.2 Factors that predict plagiarism

A comparison of the plagiarism scores between students shows that students are consistent in their plagiarism behavior. Students who plagiarize on one assignment tend to also plagiarize on other assignments. This can be demonstrated by calculating the reliability of students plagiarism scores (the length of the longest plagiarized string) across all assignments (Cronbach's alpha = 0.95, based on all 176 students from all 3 sections).

The incidence of plagiarism was significantly correlated ($r=0.5$) with the readability (defined as the FOG index) of each book chapter, suggesting that plagiarism is more likely from chapters that require a higher reading level.

3.3 Trends throughout a semester

Figure 2 shows changes in plagiarism throughout the semester (data are shown for the section that was not given plagiarism feedback and for the section that was given feedback from midterm onwards). There is considerable chapter-to-chapter variability, but the data also show a general trend for plagiarism scores to increase in as the semester progresses in the section that did not receive feedback ($r=.76, p<.05$).

The increase in plagiarism throughout the semester underlines the importance of detecting plagiarism when it occurs. Without negative feedback, plagiarism increases. This is an alarming finding, because it suggests that maintaining the *status quo* and not taking action to prevent plagiarism may in fact *promote* plagiarism.

3.4 Preventing plagiarism

Figure 2 allows a comparison between the classes that did and did not receive plagiarism feedback. The class receiving feedback during the second half of the semester does not show an increase in plagiarism as the semester progresses (this same trend is also found in the third section of the class that received plagiarism feedback throughout the entire semester).

Compared to the incidence of plagiarism without feedback (described in 3.1) the overall incidence of plagiarism was lower in the sections that received feedback: a 50-character plagiarism criterion classifies 32% of the assignments as plagiarized, and a 100-character criterion classifies 6% of the assignments as plagiarized.

These data are encouraging. They suggest that providing students with plagiarism warnings, and plagiarism feedback may be a successful strategy for training better writing habits. In their written course evaluations at the end of the semester 77% of the students claimed that they changed their writing habits based upon the plagiarism feedback.

Hopefully these improved writing habits will generalize to other classes throughout their academic careers.

4.0 Acknowledgements

I thank Dr. Lary Shaffer for providing the text files for all the chapters of his textbook, and Dr. Katherine Dunham for sharing the assignments from her section of PSY101.

5.0 References

- Shaffer, L. and Merrens, M. R. (2001). *Research stories for introductory psychology*. 1st ed. Allyn and Bacon, Massachusetts, USA.
- Wise, M. (1996) YAP3: Improved detection of similarities in computer programs and other text. In *Proceedings of SIGCSE'96*, 130-134. Philadelphia, USA.