

CAN UNDERGRADUATE STUDENTS DETERMINE WHETHER TEXT HAS BEEN PLAGIARIZED?

MIGUEL ROIG
St. John's University

In two studies undergraduate students were given an original paragraph and several rewritten versions of the paragraph, some of which were plagiarized (e.g., without a citation, superficially modified from the original) and some correctly paraphrased. Students were asked to determine whether each rewritten version had been plagiarized or correctly paraphrased. Approximately 74% of the students in both studies correctly identified the paraphrased versions. However, some of the plagiarized versions were misidentified as having been correctly paraphrased by as many as 40% to 50% of the students. Results suggest that students are often unclear as to what constitutes plagiarism and correct forms of paraphrasing.

Concern about college cheating is on the rise on many college campuses (Davis, Grover, Becker, & McGregor, 1992). Various studies have documented the extent of cheating and the types of situational and attitudinal factors that mediate academic dishonesty (see for example, Davis, Simon, Handler, & Miller, 1992; Fakouri, 1972; Haines, Diekhoff, LaBeff, & Clark, 1986; McCabe, 1992; McCabe & Trevino, 1993; Singhal, 1982).

One type of cheating that is increasingly drawing more attention is plagiarism (see for example, White, 1993). Few studies, however, have focused on plagiarism as a distinct form of cheating. One study by Hale (1987) which was designed in part to determine the incidence of reported plagiarism among college students found that 55% of the students in each of his two samples reported to have plagiarized material. His findings on the reported incidence of plagiarism are well within the 40% to 60% range of students in large scale surveys who claim to cheat (e.g., Davis, Grover et al., 1992; McCabe & Trevino, 1993).

I thank Patricia Hill for her collaboration in the first study, the results of which were presented at the 66th annual meeting of the Eastern Psychological Association, Boston, MA, March, 1995. I also thank Jennifer Miranda for her assistance in analyzing the data, and Dr. Maryellen Reardon and two anonymous reviewers for their comments on an earlier draft of this article. Requests for reprints and copies of the original and revised PKS are available from the author: Miguel Roig, Division of Social Sciences, St. John's University, 300 Howard Avenue, Staten Island, New York 10301. E-mail: ROIG@SJUVM.STJOHNS.EDU.

In another study, Karlin, Michaels, and Podlogar (1988) carried out an empirical investigation with a sample of business students designed to determine the actual incidence of a specific type of plagiarism: student-from-student copying of a 3-page library log-paper. These authors reported that a mere 3% of the students (21 out of 666) were found to have plagiarized the assignment. While Karlin et al. acknowledge that their results may have been attributable to the very specific type of plagiarism studied, their findings are somewhat surprising because they represent one of the lowest incidences of cheating in the literature. In addition, business students (i.e., accounting, finance, economics, management, and marketing majors) have been found to have some of the most tolerant attitudes toward cheating (Roig & Ballew, 1994) and have also been reported (i.e., economics majors) to be some of the most frequent cheating offenders (Moffatt, 1990).

Given that certain plagiarism practices (e.g., failure to acknowledge the original author in papers) are some of the most common forms of cheating cited by students (McCabe, 1992), the question arises as to the factors that underlie plagiarism. There are those who believe that many cases of plagiarism are accidental resulting from students' lack of knowledge regarding correct ways of citing and paraphrasing information (e.g., White, 1993; Rosnow & Rosnow, 1995). However, Hale's (1987) study has provided some evidence suggesting that plagiarism is not the result of confusion on the part of students. Kiang (1993) described a case in which 78 undergraduate students in a programming class knowingly plagiarized a class project. His analysis of the case suggested that course difficulty and competition for grades were important mediators of plagiarism. Interestingly, these factors have also been implicated in studies of traditional forms of cheating (e.g., Davis, Grover et al., 1992; Haines et al., 1986).

Unlike most common cheating behaviors (e.g., looking over another student's paper and obtaining an answer, using crib sheets during an exam), plagiarism may vary in degree from very subtle forms (e.g., moderate, but still insufficient modification of original text with acknowledgment to its author) to blatant cases (e.g., copying large portions of text with no modification and without crediting the original author). Definitions of plagiarism typically focus on the failure to acknowledge the original author of the borrowed text. However, plagiarism may occur even when the "borrower" acknowledges the author of the original material. For example, consider the situation where a writer takes a paragraph of text and changes only one or two words such as prepositions or articles, repositions the subject and predicate and includes a reference note or some other indication that the writer has credited the original author. A strict interpretation renders such "paraphrasing" as constituting a case of plagiarism, particularly if the "voice" of the original author is preserved in the rewritten version (see Campbell & Ballou, 1974).

The results of Hale's (1987) study indicate that most college

students understand traditional forms of plagiarism (i.e., failure to acknowledge the author of material being used). But, to what extent are college students aware of the more “subtle” forms of plagiarism? What criteria does the average undergraduate use to determine whether text has been properly paraphrased versus plagiarized? Does the criteria change as a function of educational experience (e.g., freshmen versus seniors) or of some other demographic factors (e.g., grade point average, Scholastic Assessment Test scores)? To shed light on these questions, I designed an instrument titled the Plagiarism Knowledge Survey (PKS). This survey consists of an original paragraph and 10 rewritten versions, 8 of which were plagiarized to various degrees and 2 of which were correctly paraphrased. The main purpose of this study was to assess college students’ understanding of plagiarism. A comprehensive grasp of the concept of plagiarism would allow students to distinguish between the various plagiarized versus correctly paraphrased versions of the original paragraph in the PKS.

Study 1

Method

Subjects. Participants in the study were 316 undergraduate students from two private colleges in the New York metropolitan area. Of those students who identified themselves by sex, 194 were women and 103 were men. Their ages ranged from 17 to 48 with an average of 19 years of age.

Materials and procedure. The PKS consists of an original paragraph taken from Zenhausern (1978) and 10 rewritten versions, all of which, except for 2, were plagiarized to various degrees from blatant to more subtle forms. The plagiarized rewritten versions were classified as such because they either lacked a proper citation and/or they were superficially paraphrased. For example, the first rewritten version of the original was simply copied word for word, without quotation marks, but a citation was added at the end of the paragraph (i.e., Zenhausern, 1978). In a later version, synonyms were substituted for some words in one of the sentences of the original paragraph but no citation was added. An attempt was made to increase the amount of paraphrasing in the rewritten plagiarized versions such that the first and second rewritten paragraphs were clearly plagiarized. Paragraphs 3 and 4 were minimally changed and Paragraphs 5, 6, and 7 were moderately but insufficiently changed to be deemed as correctly paraphrased. Some of the first 7 paragraphs lacked a citation (i.e., Zenhausern, 1978; or a footnote) which rendered them automatically plagiarized. Paragraphs 8 and 9 were sufficiently modified to have been classified as having been correctly paraphrased and both included a reference citation or footnote. Paragraph 10 which was identical to Paragraph 9 was deemed to have been plagiarized as no reference citation or footnote was included.

Four independent judges (two English professors, a Psychology professor, and an Industrial Psychologist) agreed with the plagiarism

criteria established by the author of the PKS.¹ All task instructions, demographic questions and original and rewritten versions were typed on both sides of an 8¹/₂- by 11-inch piece of paper. The original, which was boxed and clearly marked as such, appeared on both sides of the PKS and was always available to students for review when evaluating the rewritten versions.

Copies of the PKS were distributed in various introductory psychology, history, and English classes. The students' task was to read the original paragraph, examine each rewritten version, and determine whether the rewritten version was plagiarized, not plagiarized, or indeterminable whether or not it had been plagiarized.

Results

For each rewritten paragraph, the percentages of students who indicated that the paragraph had been plagiarized, not plagiarized, or indeterminable were calculated and are summarized in Table 1. Any rewritten paragraphs left blank were treated as if the student could not determine whether they had been plagiarized or correctly paraphrased.

Approximately 76% of the respondents correctly identified the two paraphrased versions of the original (80% for Paragraph 8 and 72% for Paragraph 9). However, six of the plagiarized versions of the paragraph were incorrectly identified as not having been plagiarized by approximately 50% of the students.

Table 1

| Responses of Students (<i>n</i> = 316) to Each Rewritten Version of the Original Paragraphs | | | |
|--|-------------|-----------------|------------------|
| | Plagiarized | Not Plagiarized | Cannot Determine |
| Paragraph 1 | 39% (123) | 57% (181) | 4% (12) |
| Paragraph 2 | 42% (134) | 48% (151) | 10% (31) |
| Paragraph 3 | 37% (116) | 50% (158) | 12% (42) |
| Paragraph 4 | 63% (199) | 20% (63) | 17% (54) |
| Paragraph 5 | 40% (125) | 48% (150) | 13% (41) |
| Paragraph 6 | 67% (213) | 17% (54) | 16% (49) |
| Paragraph 7 | 19% (59) | 66% (209) | 15% (48) |
| * Paragraph 8 | 8% (24) | 80% (254) | 12% (38) |
| * Paragraph 9 | 10% (32) | 72% (226) | 18% (58) |
| Paragraph 10 | 25% (80) | 54% (169) | 21% (67) |

*Indicates rewritten versions were **NOT PLAGIARIZED**

¹Although all four independent judges who examined the PKS were ultimately in agreement with the plagiarism criteria used for the PKS, one of the judges questioned whether Paragraph 7 should have also been considered as having been correctly paraphrased. A review of the *Publication Manual of the American Psychological Association* (1994) indicates that "Summarizing a passage or rearranging the order of a sentence and changing some of the words is paraphrasing" (p. 292). However, the example for correct paraphrasing offered in the *Manual* (p. 294) represents a considerable rearrangement of the original (also as suggested by Campbell & Ballou, 1974) unlike the rearrangement we used in Paragraph 7, and therefore the original criteria is maintained. The above difference of opinion may serve to illustrate the gray area that exists when judging potential cases of plagiarism.

A plagiarism knowledge (PK) score was derived for each student by adding one point for each rewritten version correctly identified, two points for each rewritten version which they could not identify, and three points for each version which they incorrectly identified as having been plagiarized. A student who would have correctly identified all 10 rewritten versions of the original would obtain a perfect score of 10. The highest possible PK score was 30.

Several analyses were conducted using PK scores as a dependent measure. The performance between students from the two colleges was compared and found not to be significantly different, $t(314) = 1.04$, $p = .30$. Similarly, no statistically significant differences in PK scores were found between men ($X = 19.21$) and women ($X = 18.69$), $t(300) = 1.37$, $p = .17$.

One might assume that knowledge about proper procedures for paraphrasing text might increase as students acquire additional writing experiences while they advance through college. To test this hypothesis differences in PK scores among the four academic years were examined. A one-way analysis of variance (ANOVA) with academic level (e.g., freshmen, sophomore) as the between subjects factor was carried out yielding a statistically significant main effect, $F(3, 274) = 4.26$, $p = .006$. Newman-Keuls showed that the PK scores of freshmen ($X = 19.13$) and juniors ($X = 19.91$) were significantly higher than those of sophomores ($X = 18.04$), $p < .05$. The mean PK score for seniors was 18.40, but it was not significantly lower than the mean PK for freshman and juniors. No other statistically significant differences were detected.

In the demographics section of the PKS students were requested to report their grade point average (GPA), Scholastic Assessment Test (SAT) scores on the verbal and math sections, and the estimated total number of papers written in college and in high school. Multiple Pearson r correlations were carried out between these data and the students' PK scores. The only statistically significant correlation occurred between GPA and PK scores, $r(139) = -.20$, $p < .009$, indicating that students with low GPAs tend to score high on plagiarism.

Discussion

The finding that over 50% of the sample incorrectly judged 6 of the plagiarized versions of the original to be correctly paraphrased is in direct contrast to Hale's (1987) conclusions. These data indicate that students appeared to be confused as to the extent to which original text needs to be modified, and about the conditions under which a citation is necessary. Consider the following pattern of responses. Paragraph 1 was copied word for word from the original, without quotation marks, and with the correct citation (i.e., Zenhausern, 1978) added at the end. Yet, 57% of the sample judged Paragraph 1 not to be plagiarized. In Paragraph 2, the order of the two sentences was simply reversed and no quotation marks nor reference citation or footnote was included, yet, 48% of the sample judged this rewritten paragraph as not having been plagiarized with an additional 10% of students not being able to determine whether the material had been plagiarized or not.

The pattern of results obtained in the present study suggests that more than half of the students in our sample were not adequately informed about the proper procedures for paraphrasing text and thus could not correctly distinguish between various types of plagiarized versus correctly paraphrased text. The significant negative correlation between GPA and PK scores indirectly adds some validity to our data, for it would be expected that being misinformed about correct writing practices contributes to poorer academic performance. The finding that freshmen had significantly higher PK scores than sophomores makes sense from the point of view that students' knowledge about writing should improve as a function of academic experience. However, the fact that juniors as a group produced the highest PK scores is not consistent with the above assumption and no adequate explanation could be found to account for this apparent anomaly.

Study 2

After the first study was completed I converted the PKS into an instructional sheet for the purpose of clarifying the differences between correct paraphrasing and subtle versus blatant instances of plagiarism. Feedback provided by some of the students during these instructional sessions seemed to support the basic interpretation of the results obtained in the first study. That is, as long as the original author is acknowledged, many students seem to believe that it is proper to take portions of text, with little or no modification, and to appropriate such text as their own writing.

Based on the informal student feedback, and in an effort to clarify the findings of the previous study, the PKS was revised as follows. Four of the paragraphs were deleted and notations, such as footnotes and author-year in parentheses, were removed from the remaining rewritten versions. For the present study, students were led to assume that the correct citation appeared in the rewritten version. In addition, the instructions were amended by asking participants, when responding to the rewritten versions, to consider whether each rewritten version had been changed sufficiently enough so as to not be classified as a case of plagiarism.

Method

Subjects. A new sample of 231 undergraduate students was recruited from the same two institutions. Of those students who indicated their sex, 116 were women and 104 were men. Their ages ranged from 17 to 46 with an average of 20 years of age.

Materials and procedure. The revised PKS contained the same original paragraph taken from Zenhausern (1978) and 6 of its rewritten versions (Paragraphs 1, 3, 5, and 9 in the original PKS were deleted). All citation information (e.g., footnote notation) was removed from the rewritten versions. Instead, students were led to assume that each

rewritten paragraph would be correctly cited according to the specific writing style used in the student's profession (e.g., MLS, APA). As in the previous study, students were asked to read the original paragraph and then each rewritten version, and to determine whether each rewritten version was correctly paraphrased, plagiarized, or indeterminable whether or not it was correctly paraphrased or plagiarized. In examining each rewritten version students were specifically asked: "How different does the rewritten, paraphrased version have to be so as to not be classified as a case of plagiarism?"

The demographics section of the PKS was also amended by adding two items asking students if they had ever plagiarized written material and if they had ever been caught plagiarizing. The PKS was distributed in various sections of introductory and advanced psychology, history, political science, and computer science courses at the same two private institutions.

Results

As with the previous study, the percentage of responses to each of the three categories was calculated. These data are shown in Table 2 together with the corresponding percentages of responses to these same paragraphs from Study 1.

Approximately 72% of the students correctly identified the two paraphrased versions of the original (82% for Paragraph 5 and 62% for Paragraph 6). This particular response pattern was very similar to that obtained in the first study.

PK scores were computed for each student using the procedure outlined in the first study. For this version of the PKS a student who correctly identified all 6 rewritten versions of the original would obtain a perfect score of 6. The highest possible PK score was 18. A set of statistical analyses similar to those conducted with data from the first study were carried out. A comparison of students' PK scores from both

Table 2

| Responses of Students ($n = 231$) to Each Rewritten Version of the Original Paragraphs | | | |
|--|-------------|-----------------|------------------|
| | Plagiarized | Not Plagiarized | Cannot Determine |
| Paragraph 1 | 73% (170) | 18% (41) | 9% (20) |
| Paragraph 2 | 37% (134) | 50% (158) | 12% (42) |
| | 57% (131) | 29% (67) | 14% (33) |
| Paragraph 3 | 63% (199) | 20% (63) | 17% (54) |
| | 62% (144) | 21% (48) | 17% (39) |
| Paragraph 4 | 67% (213) | 18% (54) | 14% (49) |
| | 19% (43) | 65% (150) | 17% (38) |
| * Paragraph 5 | 19% (59) | 66% (209) | 15% (48) |
| | 7% (17) | 82% (189) | 11% (25) |
| * Paragraph 6 | 8% (24) | 80% (254) | 12% (38) |
| | 14% (32) | 62% (144) | 24% (55) |
| | 10% (32) | 72% (226) | 18% (58) |

*Indicates rewritten versions were **NOT PLAGIARIZED**

colleges yielded a small but statistically significant difference between the two samples, $t(228) = 2.09$, $p = .04$. In addition, men's ($X = 10.45$) PK scores were significantly higher than those of women ($X = 9.55$), $t(218) = 3.37$, $p = .0007$.

A one-way ANOVA with academic level (e.g., freshmen, sophomore) as the between subjects factor was carried out on students' PK scores, but no statistically significant main effect occurred with this sample. In addition, multiple Pearson r correlations were conducted with PK scores, GPA, reported verbal and math SATs, and number of papers written in high school and in college. PK scores correlated negatively with reported GPA, $r(109) = -.22$, $p < .009$, and with math SATs, $r(109) = -.19$, $p < .02$. The correlation between verbal SATs and PK scores was also negative $r(109) = -.11$, $p = .12$, but failed to reach statistical significance.

Finally, the two additional items in the demographics section yielded the following results. Seventy-two students out of 199 (36%) who responded to the question: "Have you ever plagiarized written material" admitted to have plagiarized. Of the 199 students who responded to "Have you ever been caught plagiarizing?", only six students out of 197 (3%) responded yes.

Discussion

Unlike the original, the modified version of PKS yielded an overall response pattern that paralleled more closely the plagiarism criteria of the PKS agreed to by the independent judges. Thus, it is possible that students in the first study may have partly based their responses on the presence or absence of citation information. It is noteworthy, however, that student responses from the two rewritten versions of paragraphs that had been deemed to be correctly paraphrased yielded data that was almost identical to the responses obtained from the same paragraphs in the first study.

Perhaps the most salient feature of these data concerns the pattern of responding to Paragraph 4 (Paragraph 7 in the previous study). Although this particular rewritten version contained the original author's name (though not in the form of a citation as it lacked the year of publication) the changes themselves were superficial (see previous footnote), yet a disproportionate 65% of the students felt that this version was not plagiarized. This finding adds weight to the premise that the presence or absence of acknowledgment to the original author is perhaps the primary criteria that students use to determine whether text has been plagiarized. The degree of modification of such text seems of lesser importance to the student.

The statistically significant difference between the two samples may be the result of instructional experiences at one of the institutions placing stronger emphasis on correct writing practices. On balance, the finding that men produced higher PK scores than women is interpreted in the context of a known trend in the literature of sex differences which shows that women tend to approach cognitive tasks more carefully than men

(see Maccoby & Jacklin, 1974). That the above findings were detected with the revised version, but not with the original PKS, may represent evidence of increased sensitivity in the current version of this instrument.

The negative correlations between PK scores, GPA, and SATs suggest that knowledge of correct paraphrasing practices (lower PK scores) should be associated with higher academic achievement as represented in these self-report measures. The two modest, but statistically significant, negative correlations are also taken as evidence of the divergent validity of the PKS.

Thirty-six percent of the sample admitted to having plagiarized material, whereas only 10 students out of the total sample (4%) obtained perfect PK scores. Assuming that lack of knowledge of correct paraphrasing practices translates into plagiarism, then approximately 60% of students (in addition to the 36% who already admit to plagiarize) in the present study who claim *not* to plagiarize may have actually engaged in inadvertent plagiarism. With so many students committing plagiarism, whether intentional or inadvertent, it is surprising that only 3% admitted to have been caught plagiarizing.

General Discussion

If PK scores reflect the students' own paraphrasing practices, the present findings represent evidence for the position that the majority of students probably engage in inadvertent plagiarism. The overall picture that emerges from the results of both studies is that as long as the original author is credited and/or as long as minor modifications are made to the original, the material is generally considered to be properly paraphrased. In contrast to the findings of Karlin et al. (1988) and Hale (1987), the present findings suggest that plagiarism may be a larger problem than previously thought, and that a substantial amount of this activity may stem from ignorance, on the part of students, over the proper rules for correctly paraphrasing text.

That students lack the necessary knowledge to determine whether text has been correctly paraphrased or plagiarized is consistent with recent observations concerning the apparent decline in students' writing abilities (Shea, 1993). The decline and consequent need to strengthen writing skills in students has been recognized in our own field of psychology (see for example *Teaching of Psychology*, Special Issue, 1990). However, in addition to attempting to improve general writing skills, more attention needs to be paid to teaching students the proper skills to avoid plagiarism.

Because it is not uncommon these days for undergraduate students to be involved in all levels of scientific research, the potential for plagiarism, whether blatant or subtle, becomes a matter of serious professional concern. Given that the present results suggest that a large number of students may be committing inadvertent plagiarism, a situation is likely to arise where a relatively simple matter of academic dishonesty may translate into a more serious case of scientific misconduct.

References

- CAMPBELL, W. G., & BALLOU, S. V. (1974). *Form and style: Theses, reports, term papers* (5th ed.). Boston: Houghton Mifflin.
- DAVIS, S. F., GROVER, C. A., BECKER, A. H., & MCGREGOR, L. N. (1992). Academic dishonesty: Prevalence, determinants, techniques, and punishments. *Teaching of Psychology*, 19, 16-20.
- DAVIS, S. F., SIMON, L., HANDLER, C., & MILLER, K. (1992). *Academic dishonesty: One-time offenders or habitual criminals?* Paper presented at the annual meeting of the American Psychological Association, Washington, DC.
- FAKOURI, M. E. (1972). Achievement motivation and cheating. *Psychological Reports*, 31, 629-630.
- HAINES, V. J., DIEKHOFF, G. M., LABEFF, E. E., & CLARK, R. E. (1986). College cheating: Immature, lack of commitment, and the neutralizing attitude. *Research in Higher Education*, 25, 342-354.
- HALE, J. L. (1987). Plagiarism in the classroom. *Communication Research Reports*, 4, 66-70.
- KARLIN, M., MICHAELS, C., & PODLOGAR, S. (1988). *Research in Higher Education*, 29, 359-364.
- KIANG, N. (1993). *Plagiarism in a multicultural academic setting*. Paper presented at the ORI/AAAS Conference on Plagiarism and Theft of Ideas, Washington, DC.
- MCCABE, D. L. (1992). The influence of situational ethics on cheating among college students. *Sociological Inquiry*, 62, 365-374.
- MCCABE, D. L., & TREVINO, L. K. (1993). Academic dishonesty: Honor codes and other contextual influences. *Journal of Higher Education*, 64, 521-538.
- MACCOBY, E. E., & JACKLIN, C. N. (1974). *Psychology of sex differences*. Stanford, CA: Stanford University Press.
- MOFFATT, M. (1990). *Undergraduate cheating*. ERIC Document Reproduction Service No. ED 334921.
- Publication Manual of the American Psychological Association* (4th ed.). (1994). Washington, DC: American Psychological Association.
- ROIG, M., & BALLEW, C. (1994). Attitudes toward cheating in self and others by college students and professors. *The Psychological Record*, 44, 3-12.
- ROSNOW R. L., & ROSNOW, M. (1995). *Writing papers in psychology* (3rd ed.). Brooks/Cole: Pacific Grove, CA.
- SHEA, C. (1993). What's happened to writing skills? *The Chronicle of Higher Education*, February 3, A33-34.
- SINGHAL, A. C. (1982). Factors in students' dishonesty. *Psychological Reports*, 51, 775-780.
- TEACHING OF PSYCHOLOGY (1990). *Special issue: Psychologists teach writing*, 17.
- WHITE, E. M. (1993). Too many campuses want to sweep student plagiarism under the rug. *The Chronicle of Higher Education*, February 24, A44.
- ZENHAUSERN, R. (1978). Imagery, cerebral dominance, and style of thinking: Unified field model. *Bulletin of the Psychonomic Society*, 12, 381-384.