

Prevalence of Plagiarism among Medical Students

Lidija Bilić-Zulle,^{1,2} Vedran Frković,³ Tamara Turk³ Josip Ažman,³ Mladen Petrovečki^{1,4}

¹Department of Medical Informatics, Rijeka University School of Medicine, ²Rijeka University Hospital Center; ³Medical students, Rijeka University School of Medicine, Rijeka; and ⁴Dubrava Clinical Hospital, Zagreb, Croatia

Aim	To determine the prevalence of plagiarism among medical students in writing essays.
Methods	During two academic years, 198 second year medical students attending Medical Informatics course wrote an essay on one of four offered articles. Two of the source articles were available in an electronic form and two in printed form. Two (one electronic and one paper article) were considered less complex and the other two more complex. The essays were examined using plagiarism detection software "WCOPYFIND," which counted the number of words from matching phrases with six or more words. Plagiarism rate, expressed as the percentage of the plagiarized text, was calculated as a ratio of the absolute number of matching words and the total number of words in the essay.
Results	Only 17 (9%) of students did not plagiarize at all and 68 (34%) plagiarized less than 10% of the text. The average plagiarism rate (% of plagiarized text) was 19% (5-95% percentile=0-88). Students who were strictly warned not to plagiarize had a higher total word count in their essays than students who were not warned ($P=0.002$) but there was no difference between them in the rate of plagiarism. Students with higher grades in Medical Informatics exam plagiarized less than those with lower grades ($P=0.015$). Gender, subject source, and complexity had no influence on the plagiarism rate.
Conclusions	Plagiarism in writing essays is common among medical students. An explicit warning is not enough to deter students from plagiarism. Detection software can be used to trace and evaluate the rate of plagiarism in written student essays.

Plagiarism is the appropriation of another person's ideas, processes, results, or words without giving appropriate credit and usually claiming it to be one's own (1-5). In the scientific community, plagiarism is undoubtedly present, although it is contradictory to basic scientific principles (6,7). It is useless, meaningless, unethical, and forbidden. Plagiarism is a complex and long standing problem (1,8,9). There were well documented cases of plagiarism in the scientific community even 200 years ago (10).

The significance of plagiarism in biomedical sciences is illustrated by the existence of "plagiarism" as a regular term in the Medical Subject Headings list of the Index Medicus and a con-

stant rise in the number of published articles dealing with plagiarism (11).

It is very important to recognize plagiarism in the academic community and teach about it. It is dangerous and educationally counterproductive to ignore plagiarism among students because then they may choose plagiarism as an easier way rather than be honest, hardworking members of the academic community (5,12). Dishonest medical students may become dishonest doctors who lack medical knowledge (13). Today, the use of computers makes the plagiarism easier than ever before. Information is readily available on the Internet and other electronic media and word processing software allows a simple "copy&paste" of

the information from the source (4,5,9,14). Although computers make plagiarism easier to perform, they also make it easier to detect and quantify it (4,14). There are several software tools for detecting plagiarism, most of them using correlation techniques to discover concordances between several documents (4,14). Some plagiarism-detecting services and tools are available on the Internet, and some of them are free of charge. Internet services, e.g., Turnitin® (iParadigms, Oakland, CA, USA; www.turnitin.com) and EVE (Essay Verification Engine, CaNexus.com; www.canexus.com/eve/), allow detection of similarities between a text and documents collected from different databases or on the Internet. These services can be used only for the English language (4).

There are methods that can be used regardless of the language. The Glatt Plagiarism Service is a computer program that is not based on correlation techniques but deletes every fifth word from a student's paper and the student is asked to fill in the missing words. If a student cannot fill in 77% of the missing words from his or her paper, he or she is probably plagiarizing (4).

One of the computer programs based on correlation techniques is WCopyfind, created by Louis Bloomfield (15) which can be downloaded from the Internet (www.plagiarism.phys.virginia.edu/Wsoftware.html) free of charge. WCopyfind examines a collection of document files, extracts the text portions of these documents and looks for matching phrases between them. The software does not search the Internet for matching documents. Instead, the user needs to specify which documents should be compared. It makes this program useful for texts written in any language (15).

The aim of this study was to test the extent of plagiarism among medical students and find out if there was a difference in the plagiarism rate between female and male students. We also investigated the influence of a warning against plagiarism, source of the articles (electronic or paper copy), subject complexity, and students' examination grade.

Subjects and Methods

Subjects

The study included 198 second-year medical students of the Rijeka University School of Medicine, who attended the mandatory Medi-

cal Informatics Course. The study was conducted during two consecutive academic years, 2000/01 (n = 111) and 2001/02 (n = 87). There were 129 female and 69 male students, with the median age of 21 years (range 19-27).

Writing the Essay

During the course students had to write an essay based on one of four scientific articles. Essays had to have at least 250 words and had to be submitted in an electronic form. All students were informed that the essay was mandatory, conditional for taking the final examination in Medical Informatics. They were also told that their essay would not be graded and that the quality of the essay would not influence their final grade.

Students were told that they were expected to use offered articles as a template only and that they were expected to write their essays in their own words as much as possible. Students attending the course in the second year of the study were specifically and clearly warned that plagiarism was forbidden. They were explained that direct copying of phrases and words from the source paper would be considered plagiarism and fraud.

The four scientific papers, offered as templates, were published in the Proceedings of 4th Croatian Conference on Medical Informatics and were written in the Croatian language. Two of them were considered more complex. One more and one less complex paper were available in electronic form at the University web sites (<http://mi.medri.hr>). Students were free to choose one of the four subjects for their essay.

Analysis of Essays Using Plagiarism Detection Software

After submitting the essays, the total number of words in the body text was counted for each essay (the title and student's data were excluded). In order to use plagiarism detection software WCopyfind version 2.1 (15), Croatian characters were converted to match the international characters in both students' essays and source articles. If they were present, tables and figures were excluded from the documents due to the incapability of the software to analyze them. Program parameters were adjusted according to the author's recommendations and available published data (14,15). The shortest phrase to match was set to six words, meaning that the program would ignore matching phrases with only five or fewer words.

"Most imperfections to allow" parameter was set to "2" meaning that the program would bridge between two non-matching words and connect the pieces of a perfectly matched phrase. Punctuation, non-words (non-textual items, filenames, URLs, etc.), and numbers were ignored. After comparing the documents, the program provided an absolute number of matching words.

Plagiarism rate was expressed as the percentage of the plagiarized text and calculated as a ratio of the number of matching words and total number of words in the essay.

After the course, most students entered the final exam in Medical Informatics. The final grades from the first attempt at the examination were collected. According to the examination grades students were divided into three groups. Students who did not enter the examination for at least one year after finishing the course and those students who failed the exam on their first attempt at the examination were considered as students with poor performance. Students who got grade 2 and 3 were considered "good students" and students who obtained the grades 4 and 5 were considered "excellent students."

Statistical Analysis

Total word count in the essay, matching words count, and the plagiarism rate were analyzed for the differences according to student's gender, subject source, subject complexity, plagiarism warning, and exam grade. Kolmogorov-Smirnov test indicated that data distribution differed significantly from normality. Therefore, Mann-Whitney test was used for two group comparison and Kruskal-Wallis test for three group comparison. When the Kruskal-Wallis test was significant, post hoc comparison was performed using Mann-Whitney test with the adjustment of *P* value. Association of the numerical parameters was tested by calculating Spearman's correlation coefficient (16).

Data is expressed as median and a 5th-95th percentile range. Variables were analyzed as independent, with no adjustment. *P* values less than 0.05 was considered significant. In post hoc comparison, *P* less than 0.017 was considered significant for three independent comparisons. Computation was performed using MedCalc statistical software version 7.3.0.0 for Windows (MedCalc Inc., Mariakerke, Belgium)

Results

Only 17 (9%) students did not plagiarize at all and 68 (34%) plagiarized at a rate lower than 10% (Fig. 1). The median plagiarism rate was 19% (5-95% percentile=0-88) of the text.

Students who were strictly warned not to plagiarize had a larger average total word count in their essays compared to the others (Table 1, 553 words vs 455 words, *P*=0.002) but no difference was found according to the matching word count and the plagiarism rate.

Although the average plagiarism rate in the papers of the students who got excellent grades at Medical Informatics exam was 12%, ranging from 0% to 72% (5th and 95th percentile), they plagiarized less than those who got good (median 22% [0%-94%]) and poor (median 25%, [0%-89%]) grades. Student's gender, the subject

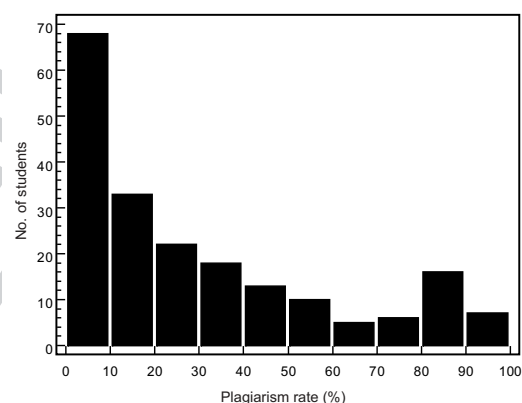


Figure 1. Distribution of students according to the plagiarism rate.

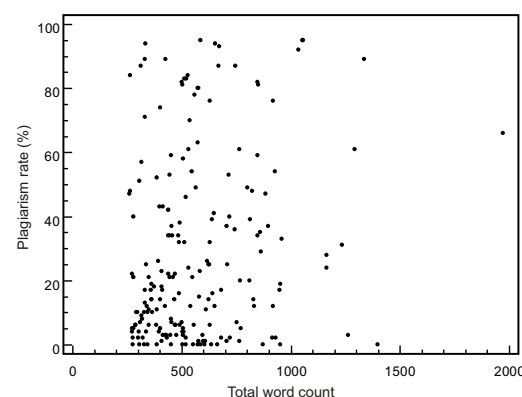


Figure 2. Correlation between total word count and plagiarism rate (Spearman's $\rho=0.19$, $p=0.009$).

Table 1. Total word count, matching word count, and plagiarism rate among medical students writing an essay from medical informatics

Variable	Median (5-95% percentile range) of					
	total word count	<i>P</i> *	matching word count	<i>P</i> *	plagiarism rate	<i>P</i> *
Gender:						
female (n=129)	510 (289-1062)	0.390	92 (0-702)	0.758	18 (0-89)	0.679
male (n=69)	504 (274-951)		112 (0-559)		19 (0-87)	
Subject source:						
hard copy (n=97)	518 (280-939)	0.629	98 (0-572)	0.671	19 (0-86)	0.847
electronic (n=101)	504 (283-1106)		92 (0-694)		17 (0-92)	
Subject complexity:						
less complex (n=153)	503 (280-1039)	0.234	70 (0-608)	0.056	17 (0-87)	0.065
more complex (n=45)	538 (285-1068)		128 (0-693)		24 (0-92)	
Warned against plagiarism:						
yes (n=87)	553 (310-1165)	0.002	105 (0-636)	0.710	21 (0-87)	0.758
no (n=111)	455 (280-948)		74 (0-650)		17 (0-89)	
Medical Informatics exam grade:						
poor (n=49)	455 (306-1041)	0.556	151 (0-701)	0.046	25 (0-89)	0.015
good (n=92)	517 (279-1057)		119 (0-647)		22 (0-94)	
excellent (n=57)	541 (288-943)		56 (0-442)†		12 (0-72)†	
Total (N=198)	507 (280-1047)		94 (0-641)		19 (0-88)	

*Mann-Whitney test for two groups and Kruskal-Wallis for three groups comparisons.

†Significantly different from other two groups.

source, and subject complexity did not influence the total word count or the plagiarism rate.

Spearman's ρ coefficient was very low ($\rho=0.19$, $P=0.009$). Although significant, low correlation indicated that the length of the text is not important for plagiarism (Figure 2).

Discussion

Our findings that only 9% of students did not plagiarize, and that all the others copied between 1% and 95% of the text are disturbing.

Braumoeller and Gaines reported that plagiarism detection software found 12% of the students' papers under suspicion of plagiarism among students of political science (17). All other available studies on plagiarism among students are based on surveys ie anonymous questionnaires reporting behavior. One such study showed that 16.5% of students from 9 different colleges and universities in the USA reported that they plagiarized whereas 50% of them reported that they thought their peers had frequently copied and pasted text from the Internet into their papers without proper citation (18). According to a survey of 3,500 students in US and Canadian universities, 23-25% acknowledged one or more "copy & paste" episodes (5). Plagiarism is more often found in "hard sciences," such as medicine, than in social sciences (5). Rennie, for example, reported that 56% of medical students reported having done or considering "copying text directly and just including the source" and 14% would do this without acknowledging the source (8).

Our data also supports the study of Hrabak et al (13), conducted at the Zagreb University School of Medicine in Croatia, which showed that 94% of the students admitted some kind of academic misconduct at least once during their studies. This is similar to 91% of students really plagiarizing in our study.

In our study, the average plagiarism rate was 19%, ranging from 0% to 88% (5th and 95th percentile).

The limitations of our study include the limitations of the plagiarism detecting software WCopyfind, because it can compare only the documents that are already available. We instructed the students to use a single source paper for their work and they were not encouraged (but also not banned) to use other sources. Thus, the analysis was performed only with respective source articles but students could have used more articles and copied from undeclared sources.

Several authors showed that easily available source text in an electronic form makes plagiarism easier and increases the rate of "copy & paste" incidents (9,12,14). We intentionally offered students two articles in electronic form, available from the university web site, and two printed in a journal, expecting that the electronic versions would be plagiarized more. However, no differences in the plagiarism rate between the students who chose the electronic version of the source article or the paper version were actually found. Students at Rijeka University School of Medicine are still not regular users of computers

and Internet tools (19). Other studies showed that the growth in the Internet plagiarism did not necessarily have to lead to an overall increase in plagiarism as yet (18).

One of the hypotheses of our study was that students would probably plagiarize more from more complex articles than from simpler texts but this was not confirmed. Furthermore, male and female students plagiarized equally. To the best of our knowledge, there was no data suggesting association of gender and plagiarism frequency.

It could also be expected that a clear warning against plagiarism would affect the rate of plagiarism among students but did not influence the rate of plagiarism among students. Our findings are consistent with the finding that not even the strongest warning appears to have an effect (17). The fact that warnings do not have any effect could be attributed to the fact that students are likely to take the risk and plagiarize when they do not expect to get caught. However, our results revealed that students who were warned submitted essays containing significantly more words than the others. Regarding the strict warning and more detailed guidelines for essay writing, we suspect that students who were warned took the task more seriously but not seriously enough to stop plagiarizing. When plagiarism detecting software becomes more widely accepted at universities the chances of getting caught will dramatically increase and the willingness of students to plagiarize will probably decrease (17).

The only significant parameter that influenced the plagiarism rate was the exam grade (Table 1). The grade is the best possible indicator of students' knowledge (19) and our study showed that it might indicate students' honesty as well. Students who obtained better exam grades plagiarized significantly less than those who got lower grades or failed the exam. The fact that has to be considered is that the essays were not graded and did not influence the final exam grade, and this was known by students at the time of writing. This could be one explanation why the majority of students did not take the task seriously.

The lack of correlation between the total word count and the plagiarism rate showed that plagiarism rate is not affected by the length of the essay.

Studies on plagiarism also reveal that students are often confused and do not exactly know what plagiarism is (8,12,18,20). This raises many issues such as copying text directly and just including the source in the reference list, which is regularly not considered to be plagiarism (8). Teachers at universities should be well aware of plagiarism among students and educate students on academic integrity, so that they produce bright intellectuals and not "copy & paste" technicians.

Universities should discourage plagiarism by establishing rules analogous to those found in editorial policies of some biomedical journals (20-23).

Acknowledgement

This study is part of the scientific project "Prevalence, features and attitudes on plagiarism in biomedical science" (No. 0062044) supported by the Ministry of Science, Education and Sports, Republic of Croatia.

References

1. Gilbert FJ, Denison AR. Research misconduct. *Clin Radiol.* 2003;58:499-504.
2. Gonsalus CK. Scientific misconduct. Policy issues. In: Jones AH, McLellan F, editors. *Ethical issues in biomedical publication.* Baltimore: The Johns Hopkins University Press; 2000. p. 223-49.
3. Martin B. Plagiarism by university students: the problem and some proposals. Available from: <http://www.uow.edu.au/arts/sts/bmartin/pubs/92tert.html>. Accessed: August 28, 2004.
4. Austin MJ, Brown LD. Internet plagiarism: developing strategies to curb student academic dishonesty. *The Internet and Higher Education.* 1999;2:21-33.
5. Das SK. Plagiarism in higher education: is there a remedy? *Scientist.* 2003;17:8. Available from: http://www.the-scientist.com/yr2003/oct/opinion_031020.html. Accessed: June 10, 2004.
6. Marušić M. Scientific integrity [in Croatian]. In: Marušić M. editor. *Uvod u znanstveni rad u medicini.* Zagreb: Medicinska naklada; 2004. p. 191-200.
7. Institute of Medicine Committee on Assessing Integrity in Research Environments. *Integrity in scientific research: creating an environment that promotes responsible conduct.* Washington (DC): The National Academies Press; 2002. p. 167-77.
8. Rennie SC, Crosby JR. Are "tomorrow's doctors" honest? Questionnaire study exploring medical students' attitudes and reported behaviour on academic misconduct. *BMJ.* 2001;322:274-5.
9. De Voss D, Rosati AC. "It wasn't me, was it?" Plagiarism and the web. *Computers and Composition.* 2002;19: 191-203.
10. Hansen TW. Neonatal jaundice and scientific fraud in 1804. *Acta Paediatr.* 2002;91:1135-8.
11. Clouthier SG. Institutionalized plagiarism. *Scientist.* 2004;18:8. Available from: <http://www.the-scientist.co>

- m/yr2004/aug/opinion_040802.html*. Accessed: June 10, 2004.
- 12 Walker JR. Copyrights and conversations: intellectual property in the classroom. *Computers and Composition*. 1998;15:243-51.
- 13 Hrabak M, Vujaklija A, Vodopivec I, Hren D, Marušić M, Marušić A. Academic misconduct among medical students in a post-communist country. *Med Educ*. 2004;38:276-85.
- 14 Grover D. Plagiarism and the internet: the use of correlation techniques to detect plagiarism. *Computer Law and Security Report*. 2003;19:36-8.
- 15 Bloomfield L. WCopyfind 2.1 instructions. Available from: <http://www.plagiarism.phys.virginia.edu/WCopyfind%202.1.html>. Accessed: October 1, 2002.
- 16 Dawson B, Trapp RG. *Basic and clinical biostatistics*. 3rd ed. New York: Lange Medical Books; 2001.
- 17 Braumoeller BF, Gaines BJ. Actions do speak louder than words: deterring plagiarism with the use of plagiarism-detection software. *Political Science and Politics Online*. Available from: <http://www.apsanet.org/PS/dc01/braumoeller.cfm>. Accessed: April 4, 2004.
- 18 Kellogg AP. Students plagiarize less than many think, a new study finds. *The chronicle of higher education, Information technology*. Available from: <http://www.chronicle.com/free/2002/02/2002020101t.htm>. Accessed: April 1, 2004.
- 19 Petrovečki M, Rahelić D, Bilić-Zulle L, Jeleč V. Factors influencing medical informatics examination grade - can biorhythm, astrological signs, seasonal aspect, or bad statistics predict outcome? *Croat Med J*. 2003;44:69-74.
- 20 Guide to plagiarism and cyber-plagiarism. University of Alberta Libraries. Available from: <http://www.library.ualberta.ca/guides/plagiarism>. Accessed: June 4, 2004.
- 21 Petrovečki M, Scheetz MD. Croatian Medical Journal introduces culture, control, and the study of research integrity. *Croat Med J*. 2001;42:7-13.
- 22 Guidelines for Authors: Editorial Policy. *Croat Med J*. 2004;45:119-26.
- 23 Bloomfield L. The importance of writing. Available from: <http://www.plagiarism.phys.virginia.edu/essays/The%20Importance%20of%20Writing.html>. Accessed: August 28, 2004.

Received: September 17, 2004

Accepted: December 21, 2004

Correspondence to:

Lidija Bilić-Zulle

Department of Computer Science

Rijeka University School of Medicine

Braće Branchetta 20

51000 Rijeka, Croatia

lbilicz@medri.hr