As Computer Coding Classes Swell, So Does Cheating

By JESS BIDGOOD and JEREMY B. MERRILL, New York Times, May 29, 2017

A Computer Science 50 course at Harvard in 2013. Last fall, more than 60 computer science students were referred to the university's honor council, which investigates cheating allegations.

For a simple assignment, writing code that would allow one computer to visit web pages located on another, two students in H. E. Dunsmore's class at Purdue University turned in nearly 100 identical lines of code. Was it a fluke? Or had they cheated?

As he looked over their work, Mr. Dunsmore, a veteran computer science professor, saw what he called the smoking gun:

boolean done = true;

while (!done) {

Because they had written **!done** — the exclamation point means "not" — the program translated it as "not true," which made the program ignore the code that followed, causing it to fail. In a class of about 450, they were the only ones who made that fatal mistake.

"This is pretty strong evidence that one had copied the other," Mr. Dunsmore said. "They later both confessed to collusion."

College students have flooded into computer science courses across the country, recognizing them as an entree to coveted jobs at companies like Facebook and Google, not to mention the big prize: a start-up worth millions.

Perhaps this is a naive question coming from someone who knows little about computer science?Why aren't there new problems given to students...

The exploding interest in these courses, though, has coincided with an undesirable side effect: a spate of high-tech collegiate plagiarism. Students have been caught borrowing computer code from their friends or cribbing it from the internet.

"There's a lot of discussion about it, both inside a department as well as across the field," said Randy H. Katz, a professor in the electrical engineering and computer science department at the University of California, Berkeley, who discovered in one year that about 100 of his roughly 700 students in one class had violated the course policy on collaborating or copying code.

Computer science professors are now delivering stern warnings at the start of each course, and, like colleagues in other subjects, deploy software to flag plagiarism. They have unearthed numerous examples of suspected cheating.

At Brown University, more than half the 49 allegations of academic code violations last year involved cheating in computer science.

At Stanford, the alma mater of the founders of Google, Snapchat and countless other internet wonders, as many as 20 percent of the students in one 2015 computer science course were flagged for possible cheating.

And at Harvard, where Computer Science 50 is practically its own brand, with T-shirts, slickly produced videos and an online audience of thousands, the class distinguished itself last fall in a more dubious way: According to The Harvard Crimson, more than 60 students were referred to the university's honor council, a committee that reviews allegations of academic dishonesty, such as plagiarism, and violations of the honor code.

In interviews, professors and students said the causes were not hard to pin down.

To some students drawn to the classes, coding does not come easily. The coursework can be time-consuming. Troves of code online, on sites like GitHub, may have answers to the very assignment the student is wrestling with, posted by someone who previously took the course.

"You've got kids who were struggling with spending a third of their time on their problem sets with the option to copy from the internet," said Jackson Wagner, who took the Harvard course in 2015 and was not accused of copying. "That's the reason why people cheat."

Complicating matters is the collaborative ethos among programmers, which encourages code-sharing in ways that might not be acceptable in a class. Professors also frequently allow students to discuss problems among themselves, but not to share actual code, a policy that some students say creates confusion about what constitutes cheating.

The executive committee at Yale, where five students were accused of copying code last fall, though a couple had charges withdrawn, acknowledged as much.

Prominent cheating scandals have cropped up at Harvard in other subjects, including in a government class. Credit Charlie Mahoney for The New York Times

"It is often such a complex task to read these guidelines," wrote the committee chairman, Paul North, a professor of German, "that the code to be written seems simple in comparison."

Penalties can be as small as a zero on an assignment, which is what the Purdue students received; they were also docked a letter in the course's final grade. (Dr. Dunsmore credits awareness of such penalties with a reduction in cheating in his courses over the last decade.) More serious or repeated infractions can lead to failing the course, or even suspension.

At some colleges, computer science courses have become a leading source of academic dishonesty complaints. But many computer science professors reject any notion that their students are more prone to cheating than others. Cheating scandals have cropped up in other subjects in recent years, including in a government class at Harvard and in a Dartmouth class on sports, religion and ethics.

 "It's mostly that a lot of computer science professors actually check," said Alex Aiken, who heads the computer science department at Stanford.

Though coding is a foreign language to most people, the principles of plagiarism are the same as with papers written in English.

Sometimes it is acceptable for lines of code to be identical, if the code is performing a routine task or one that cannot be done a different way. But other times it is a red flag.

Plagiarizers might try to cover their tracks by replacing words with synonyms, like **0** instead of **NULL**, or replacing the original author's names for variables with their own. In a hypothetical example of cheating he sometimes sees, Thomas Doeppner, the vice chairman of computer science at Brown, provided a snippet of coding for a program that finds a path through a maze:

struct visit_list node;

node.row = row;

node.col = col;

node.next = NULL;

Someone copying the code, he said, might render it this way:

struct rooms_visited room;

room.r = r;

room.c = c;

room.link = 0;

Usually, anti-cheating software can uncover these tricks. One, developed by Dr. Aiken, is called MOSS, for Measure of Software Similarity. Another program, developed by a British company, Codio, monitors students' keystrokes; a sudden burst raises questions about where the code came from.

At Harvard, David J. Malan, the CS50 professor, introduced a "regret clause," letting students who cheat and admit it within 72 hours receive an unsatisfactory or failing grade on the assignment, and avoid further discipline — unless they do it again.

But in last fall's CS50, an unusually large number of cheating allegations — involving more than 60 out of 655 students, The Crimson reported — were reported by Professor Malan or his staff to Harvard's academic honor council.

Professor Malan said one reason for the large number of cases was that instructors were scrutinizing students' work more carefully than before. He also said that since students could use the regret clause, instructors felt more comfortable going to the honor council when students had passed up that chance.

He referred questions about the outcomes of the cheating complaints to a university spokeswoman, who said she could not discuss the cases because of student privacy.

Two Harvard students who said they were familiar with some of the violations, and who wanted to be anonymous for fear of repercussions from the university, described one of the more blatant examples: A student submitted code using a computer language, PHP, that had been taught in a previous year, but not last fall.

Some students suggested that in other cases there was a fine line between reasonable sharing and improper copying, particularly because the professor encourages students to discuss problems with one another.

"There is a gray area as to how much help one can get," said Javier Cuan-Martinez, the president of the Harvard Computer Society, who took the course two years ago.

Professor Malan, in an email, said the guidelines in the course syllabus were clear. One rule says that when asking for help, you may have your code viewed by others, but you may not view theirs. "Whenever a line is actually crossed," Professor Malan wrote, "it's often that one."