

# Metrics for Student Classroom Engagement and Correlation to Software Assignment Plagiarism

William Allen  
Rensselaer Polytechnic Institute  
allenw@rpi.edu

Jenay Barela  
Rensselaer Polytechnic Institute  
barelj@rpi.edu

Shelly Belsky  
Rensselaer Polytechnic Institute  
belsks@rpi.edu

Matthew Peveler  
PopSQL, Inc.  
matt.peveler@popsql.com

Ben Kelly  
Rensselaer Polytechnic Institute  
kellyb3@rpi.edu

Barbara Cutler  
Rensselaer Polytechnic Institute  
cutler@cs.rpi.edu

## ABSTRACT

Code plagiarism is an unfortunate problem which often plagues introductory level programming courses. In this poster, we study the relationship between plagiarism and student engagement using Submittity, our open-source course management, homework submission, and autograding system. We measure student participation through responses to live-lecture polls, posts in the course discussion forum, and daily and weekly office hours attendance patterns. While our assignments have strict deadlines, students are empowered to adapt deadlines by using penalty-free “late day” deadline extensions. Students who regularly participate in lecture polls are rewarded with additional late days. Submittity includes an open-source plagiarism detection module, Lichen, that detects software similarity that suggests academic dishonesty. Our implementation can be configured to examine every version of the assignment submitted by students and search for similarities between any version for any other student. We leverage the advantages of having a single, unified platform that combines submission timestamp data, course participation metrics, and code similarity data. Having many types of data stored in the same place allows us to effectively identify at-risk students earlier. We will present an exploration and analysis of data from an introductory level computer science course.

## CCS CONCEPTS

• **Applied computing** → **Learning management systems**.

## KEYWORDS

Plagiarism, Learning Management, Lecture Polling, Office Hours

## ACM Reference Format:

William Allen, Shelly Belsky, Ben Kelly, Jenay Barela, Matthew Peveler, and Barbara Cutler. 2022. Metrics for Student Classroom Engagement and Correlation to Software Assignment Plagiarism. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 2 (SIGCSE 2022)*, March 3–5, 2022, Providence, RI, USA. ACM, New York, NY, USA, 1 page. <https://doi.org/10.1145/3478432.3499133>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).  
SIGCSE 2022, March 3–5, 2022, Providence, RI, USA  
© 2022 Copyright held by the owner/author(s).  
ACM ISBN 978-1-4503-9071-2/22/03.  
<https://doi.org/10.1145/3478432.3499133>

## 1 BACKGROUND

A significant feature of Submittity is students’ ability to submit multiple versions of their work, and receive immediate feedback via automated grading. The platform includes a penalty-free assignment extension policy, “late days”, which aims to reduce student stress, and allow students to solidify their understanding of the material, decreasing the pressure of a fixed deadline. Students are given an initial number of late days in the beginning of a semester, and may be given the opportunity to earn more over time.

Many institutions rely upon existing plagiarism detection techniques and platforms [1]. Prior work studied the degree of plagiarism as the code evolves over time, over multiple submissions [2]. In this work, we use the results of these similarity analyses to understand the factors which contribute to the occurrence of code plagiarism over the submission period for a given assignment.

We hypothesize that students who are more actively and regularly engaged in classroom activities may have a lower incidence of software similarity, and that students who have access to late days will be less likely to plagiarize. We also discuss the prevalence of students using plagiarized work with the intention to replace it with their own later on when given rewards for submitting early.

## 2 METHODS & RESULTS

We study student activity data collected between the release date and final due date for assignments, and their correlation to the maximum similarity score for each submission over time. We have collected data on student activity in classes, which includes utilization of resources such as office hours, the class discussion forum, attendance of lectures, and in-class poll responses.

On the poster we will present charts showing the degree of correlation between students who have exhausted their supply of late days and the incidence of plagiarism on those assignments. We also analyze to what extent students who earn extra late days, and thus extra deadline flexibility, have a decreased incidence of plagiarism. Additionally, we examine temporal patterns in plagiarism between student submission versions as the final deadline approaches.

## REFERENCES

- [1] Saul Schleimer, Daniel S. Wilkerson, and Alex Aiken. 2003. Winnowing: Local Algorithms for Document Fingerprinting. In *Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data* (San Diego, California). Association for Computing Machinery, New York, NY, USA, 76–85.
- [2] Lisa Yan, Nick McKeown, Mehran Sahami, and Chris Piech. 2018. TMOSS: Using Intermediate Assignment Work to Understand Excessive Collaboration in Large Classes. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education* (Baltimore, Maryland, USA) (SIGCSE ’18). Association for Computing Machinery, New York, NY, USA, 110–115.