

Nudging Students Toward Better Software Engineering Behaviors

Chris Brown, Chris Parnin
 Department of Computer Science
 North Carolina State University
 Raleigh, NC, USA
 dcbrow10@ncsu.edu, cjparnin@ncsu.edu

Abstract—Student experiences in large undergraduate Computer Science courses are increasingly impacted by automated systems. Bots, or agents of software automation, are useful for efficiently grading and generating feedback. Current efforts at automation in CS education focus on supporting instructional tasks, but do not address student struggles due to poor behaviors, such as procrastination. In this paper, we explore using bots to improve the software engineering behaviors of students using *developer recommendation choice architectures*, a framework incorporating behavioral science concepts in recommendations to improve the actions of programmers. We implemented this framework in `class-bot`, a novel system designed to *nudge* students to make better choices while working on programming assignments. This work presents a preliminary evaluation integrating this tool in an introductory programming course. Our results show that `class-bot` is beneficial for improving student development behaviors increasing code quality and productivity.

I. INTRODUCTION

Enrollment in undergraduate Computer Science (CS) courses is growing rapidly, and these classes are quickly evolving to accommodate the significant increase of students [18]. To handle the large influx of students, researchers and practitioners have developed a variety of bots to automate instructional tasks. For instance, tools such as `CoderAssist` [17], `Web-CAT` [9], and `AutoGradr`¹ are useful for grading programs and providing feedback on students’ code. Furthermore, Wilcox shows using automated tools to facilitate student learning in programming courses is valuable for improving performance, increasing submissions, and saving instructional time [30].

Despite increasing enrollment in CS courses, research shows the dropout rate in programming classes is also increasing rapidly, especially among first and second year students [3]. Studies suggest the inability to maintain students in CS will lead to a “crisis” in the software industry with necessary computing-related jobs going unfilled.² Consequently, underrepresented minorities and female students disproportionately account for those underperforming in early programming courses and dropping out of the CS major [21], [2], leading to a lack of diversity in industry.³

Automated systems are beneficial for providing feedback efficiently, however they lack the ability to support students. The goal of this work is to explore using bots to retain students in CS. While not solely due to a lack of effort, Beaubouef and Mason suggest one reason for high withdrawal in programming courses is poor behavior on coding assignments [3]. Decision-making is an important skill in software engineering,⁴ however students in programming courses frequently make poor choices and adopt bad programming behaviors when writing code for projects. For example, students often ignore software development processes, which leads to increased frustration, lower grades, and eventually abandoning CS [3].

Our prior work proposes using *nudge theory* to improve the reception of automated recommendations to developers from bots [6], [7]. Nudge theory is a behavioral science concept for improving human behavior by influencing the environment surrounding decisions, or choice architecture, without providing incentives or banning alternative choices [26]. We introduced *developer recommendation choice architectures*, a novel framework for designing recommender bots to nudge software engineers towards better development practices [7]. In this work, we explore implementing this framework in an innovative system to improve the software engineering behaviors of students on programming projects.

To discover the impact of this approach on student behavior, we performed a study implementing `class-bot`, a system that utilizes *developer recommendation choice architectures* to recommend beneficial software engineering behaviors to students on projects for a university-level introductory programming course. We evaluated the effectiveness of this system by examining the code quality of projects and productivity of students. Our results suggest automated nudges with *developer recommendation choice architectures* improved student behavior by increasing performance on assignments and encouraging students to make more substantial changes to their code and start work earlier. The contributions of this research are a novel bot to recommend software engineering practices to students and a preliminary evaluation examining the impact of *developer recommendation choice architectures* on improving the software engineering behaviors of programmers.

¹<https://autograder.com>

²<https://www.itextico.com/blog/how-to-handle-the-crisis-of-software-developer-shortage-in-the-u.s>

³<https://news.gallup.com/reports/196331/diversity-gaps-computer-science.aspx>

⁴<https://hackernoon.com/decision-making-the-most-undervalued-skill-in-software-engineering-f9b8e5835ca6/>

II. BACKGROUND

A. Nudge Theory

Behavioral science research suggests nudges are effective for improving human decision-making. For example, studies show placing healthy options near the front of a cafeteria encourages students to purchase and consume healthier options [13]. In this case, students are not rewarded for making healthy decisions nor banned choosing from junk food. Furthermore, integrating nudge theory in digital choice environments is known as *digital nudging* [29]. Prior work suggests digital nudges can encourage better privacy and security decisions of users online [1] and improve student learning and motivation [22]. We aim to explore using digital nudges to improve the programming behaviors of students on coding assignments to improve their code quality, productivity, and ultimately increase retention among early CS students.

Nudges are useful for improving human behavior because of their ability to influence the context and environment surrounding decisions, or *choice architecture* [27]. Brown and colleagues propose *developer recommendation choice architectures*, a framework that applies concepts from nudge theory into automated bots to create effective developer recommendations [7]. We incorporate this framework into `class-bot` to make recommendations encouraging students to adopt better programming behaviors while working on projects. Our goal is to discover the impact of *developer recommendation choice architectures* on the decision-making and behavior of students by analyzing code quality and development productivity on programming assignments in introductory CS classes.

B. Computer Science Education

CS education researchers have evaluated a variety of systems useful for automating instructional tasks. For example, Kaleeswaran et al. found CoderAssist grades C programs and provides feedback to students on their code in an average of 1.6 seconds [17]. Prior work also outlines strategies for incorporating automated grading tools on coding assignments [31]. Moreover, Heckman suggests incorporating real development tools in courses can improve software engineering skills [14] while Hu shows integrating GitHub pull request bots on project repositories provides fast and useful feedback to students [16]. In this work, we seek to use bots as a means to enhance development behaviors in introductory programming courses.

Researchers have also explored improving student behavior in CS courses. For example, Edwards and colleagues summarize effective and ineffective student behaviors on programming assignments [10]. Studies also show that non-automated approaches, such as in-class labs [15], agile practices [23], pair learning [32], and other instructional techniques improve student behavior and learning in undergraduate CS courses. Additionally, prediction models [12] and automated tools such as DevEventTracker [19] are useful for observing student programming habits to predict bad behaviors, such as procrastination. This research explores using automated notifications from bots to improve the decision-making of students while developing code on programming assignments.

III. CLASS-BOT

The `class-bot` system nudges students by automatically generating and updating GitHub issues on project repositories (see Figure 1). We implemented our bot using the GitHub issue tracker because this system is useful for managing bugs, enhancements, and feedback on repositories,⁵ and research suggests they are useful for making recommendations to developers [4]. To improve student programming behaviors, `class-bot` encouraged them to complete the *software development process*, or set of activities necessary to program software application. Beaubouef and Mason suggest students' failure to follow development processes factors into the high attrition rate in early programming courses, noting the typical student method "*minimally includes the processes of analysis, design, coding, testing, and documentation...Unsuccessful students often want to skip analysis and design and begin typing in code immediately.*" [3, p. 105]

Each `class-bot` issue contains sections for each software development process phase (i.e. **Rq** in Figure 1 represents the Requirements phase) and listing relevant project rubric items for that phase. Specific tasks differed based on the assignment. In general: Requirements (Rq) focuses on understanding project specifications and functionality; Design (Ds) relates to project structure; Implementation (Im) centers on the code; Testing ensures students added passing unit tests (Ut) and functional test cases (St); and Deployment (Dp) verifies the repository is ready for submission and grading based on submission instructions. We compared our system to a baseline approach using an online rubric with a similar organization.

The `class-bot` issue updates fit the definition of a nudge because they do not provide rewards for completing tasks nor prevent students from ignoring items. To improve student programming behaviors, we designed `class-bot` using the three principles of *developer recommendation choice architectures*: *Actionability*, *Feedback*, and *Locality* [7]. Here, we explain how `class-bot` incorporates each principle to recommend better development process behaviors to students.

a) *Actionability*: Actionability involves reducing user effort by automating tasks to facilitate the adoption of useful behaviors. We implemented `class-bot` to incorporate actionability by automatically analyzing repositories to determine if development process tasks are completed and update GitHub issues based on code changes. However, the baseline approach requires students to manually seek information online.

b) *Feedback*: Feedback refers to the clarity of information provided to users. Our system provides a simple feedback mechanism, displaying a red x (✗) if the requirements for an item are not met and a green check mark (✓) if the task is completed. For instance, in the Deployment phase of the `class-bot` example in Figure 1, the repository contains a `.gitignore` file but the code does not compile. With the baseline approach, students are forced to check if project expectations are met themselves.

⁵<https://help.github.com/en/articles/about-issues>



class-bot commented on Jul 16, 2020 • edited by dcbrow10 •

Hi! This bot provides daily updates tracking your progress on the software process requirements for this assignment. If you have any questions or problems, feel free to leave a comment below or email the instructor.

Rq

ProjectSpecification.pdf: ✗

Updated README: ✓

Compilation and execution steps: ✗

Ds

Project structure: ✗

Im

Added source code: ✗

St

Ut

System

Black Box Test Plan: ✗

Unit

Added unit tests: ✗

Unit tests pass: ✗

Dp

Academic Integrity Contract: ✗

.gitignore file: ✓

No bin folder and class files: ✓

Code compiles: ✗

Tests compile: ✗

Fig. 1. Example class-bot recommendation

c) *Locality*: Locality focuses on the setting of a recommendation, or when and where automated interventions occur. For *spatial* locality, class-bot recommendations are located in an issue on the repository situated with the project. For *temporal* locality, or notification timing, class-bot provides automated daily updates on software development processes based on students' code contributions. On the other hand, students are forced to search for information in an ad hoc manner at a separate location from their repository using the online rubric.

IV. METHODOLOGY

A. Experiment Design

1) *Participants*: We integrated class-bot in an introductory Java programming course. All participants were undergraduate students with varying majors and levels of programming experience. For consistency in our data, we eliminated students who eventually dropped the class. Overall, we analyzed the behavior of 35 out of the initial 42 enrolled students. We use five phases to define the software development process: *Requirements*, *Design*, *Implementation*, *Test*, and *Deployment*, as presented to students in the course curriculum.

2) *Projects*: The course consisted of seven programming assignments, six projects and a final comprehensive exercise. Projects 3-5 made up the control group to avoid beginning assignments, and class-bot was introduced on the final two assignments. All projects were submitted to individual GitHub repositories. In total, we analyzed a 151 project repositories.

B. Data Collection

To determine the impact of class-bot on student behavior, we mined project repositories to observe the quality of their work and development productivity. Data analysis and collection lasted over the course of approximately six weeks.

1) *Quality*: We evaluated quality by observing the *grades* and number of *points deducted* for project submissions. In the course, project grades were determined using realistic industry code quality metrics such as passing unit tests, functional test cases, and Checkstyle⁶ static analysis tool warnings. Additionally, students who failed to meet certain project requirements had additional points subtracted from their grade. For instance, submitting an assignment within 24 hours after the deadline resulted in a -10% late penalty. We aim to discover if class-bot impacts student behavior by improving the quality of their projects.

2) *Productivity*: We measured productivity by observing the total *number of commits*, *code churn*, time until the *first commit*, and time of the *last commit*. GitHub commits record specific changes to project files,⁷ and prior work explores using repository commits to predict student performance [25] and gamify contributions to projects [24]. Research also suggests code churn, or the number of lines added, deleted, or modified in commits, is useful for measuring developer effort and code change impact [20]. Moreover, prior work shows students who start assignments earlier receive significantly higher grades while those who procrastinate often perform worse [10]. To further study productivity, we measured the time until the first commit on repositories and the amount of time between the last commit and the assignment deadline.

We use these quality and productivity metrics to investigate the impact of bots incorporating *developer recommendation choice architectures* on the software engineering behaviors of students working on programming projects.

⁶<https://checkstyle.sourceforge.io/>

⁷<https://docs.github.com/en/desktop/contributing-and-collaborating-using-github-desktop/committing-and-reviewing-changes-to-your-project#about-commits>

TABLE I
QUALITY RESULTS

	Nudge?	Mean	Median	p-value
Grade***	No	74.29	87.66	-
	Yes	76.89	95	0.0097***
Deductions	No	-20.71	-5	-
	Yes	-9.43	0	0.0672

Quality metrics for projects with and without `class-bot`
*** denotes statistically significant results ($p\text{-value} < 0.05$)

TABLE II
PRODUCTIVITY RESULTS

	Nudge?	Mean	Median	p-value
Commits	No	9.84	7	-
	Yes	12.64	9	0.1646
Code Churn***	No	205.03	4	-
	Yes	1101.57	11	0.0348***
First Commit*** (days)	No	8.32	7.41	-
	Yes	1.99	5.94	< 0.0001***
Last Commit (hours)	No	-21.72	-1.60	-
	Yes	-9.67	-2.47	0.7909

Productivity metrics for projects with and without `class-bot`
*** denotes statistically significant results ($p\text{-value} < 0.05$)

V. RESULTS

The Mann-Whitney-Wilcoxon test ($\alpha = .05$) was used to compare project quality and student productivity metrics for assignments with and without nudges from `class-bot`.

A. Project Quality

To observe the impact of `class-bot` on code quality, we analyzed the grade and points deducted on student assignments. These findings are presented in Table I. We discovered students received significantly higher scores on projects with automated nudges, indicating notifications from our system improved the quality of programming assignments. Additionally, while there was not a significant difference, we noticed projects without automated nudges lost an average of 11 more points than those with `class-bot` interventions.

B. Student Productivity

To discover the impact of `class-bot` on student productivity, we analyzed total commits, code churn, first commit time, and last commit time. Our results in Table II show nudges improved the productivity of students by significantly increasing the number of lines of code modified in commits and encouraging students to start development on programming assignments earlier. Additionally, we found that on average projects with automated nudges had more commits to the repositories and projects were submitted earlier with `class-bot` notifications.

VI. DISCUSSION

Our preliminary results suggest bots incorporating *developer recommendation choice architectures* can improve development behavior. By encouraging students to follow software engineering processes, we found `class-bot` improved project quality and student productivity by boosting grades, increasing

code churn, and preventing procrastination. We believe these automated nudges are a step towards reducing the attrition rate in CS. Furthermore, these results provide implications for software engineers in industry where developers also ignore development processes and underestimate the time and effort required to complete coding tasks [5]. Despite the advantages of bots for automating programming tasks, developers find recommendations from bots ineffective and intrusive [6]. To gain insight for improving bots to effectively recommend beneficial programming behaviors, we analyzed feedback from students and found increased project validation as well as more frequent updates as reported improvements for the `class-bot` system.

a) *Validation Frequency*: Even though they started assignments approximately six days earlier with `class-bot`, students still reported waiting until the end of the project to validate their work met project expectations. For example, one student noted “*I didn’t really check them until the final day*” and another mentioned “*I checked it once at the end to make sure everything was correct but thats it*”. This indicates a need to encourage students to validate their code more frequently, which research shows improves code quality [28]. `class-bot` provides automated updates for students to passively check their progress, however future systems can be improved by incorporating interactive mechanisms, such as checklists [11], to encourage students to validate their code more frequently during the development of their projects.

b) *Update Frequency*: Students also desired more frequent updates from `class-bot`, saying they “*would like it more if it could be updated more often, or maybe later in the day*” and “*the class bot didn’t update frequently enough*”. While `class-bot` incorporated *temporal locality* by updating issues daily, students desired more frequent feedback to support them in their work. Prior work in software engineering shows that presenting recommendations to developers at “diff time” during code reviews encouraged developers to fix more bugs compared to less frequent notifications, i.e. after overnight builds [8]. Based on this feedback for `class-bot`, future systems can be improved by providing frequent notifications based on student actions, such as immediately after commits, to provide convenient and persistent recommendations for useful software engineering behaviors.

VII. CONCLUSION

This work explores using bots to improve student software engineering behaviors in introductory programming courses. We introduce `class-bot`, a novel system that incorporates *developer recommendation choice architectures* to create automated recommendations with clear feedback in a convenient setting to encourage students to adhere to software development process phases. Our results show this system significantly improved code quality and student productivity by raising grades, increasing coding activity, and encouraging students to start earlier. We conclude by providing implications for improving `class-bot` and designing future automated systems to recommend beneficial behaviors to programmers.

REFERENCES

- [1] A. Acquisti, I. Adjerd, R. Balebako, L. Brandimarte, L. F. Cranor, S. Komanduri, P. G. Leon, N. Sadeh, F. Schaub, M. Sleeper, Y. Wang, and S. Wilson. Nudges for privacy and security: Understanding and assisting users' choices online. *ACM Comput. Surv.*, 50(3), Aug. 2017.
- [2] A. Baer and A. DeOrio. A longitudinal view of gender balance in a large computer science program. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education, SIGCSE '20*, page 23–29, New York, NY, USA, 2020. Association for Computing Machinery.
- [3] T. Beaubouef and J. Mason. Why the high attrition rate for computer science students: some thoughts and observations. *ACM SIGCSE Bulletin*, 37(2):103–106, 2005.
- [4] T. F. Bissyandé, D. Lo, L. Jiang, L. Réveillère, J. Klein, and Y. L. Traon. Got issues? who cares about it? a large scale investigation of issue trackers from github. In *2013 IEEE 24th International Symposium on Software Reliability Engineering (ISSRE)*, pages 188–197, Nov 2013.
- [5] B. W. Boehm. Software engineering economics. *IEEE Transactions on Software Engineering*, SE-10(1):4–21, 1984.
- [6] C. Brown and C. Parnin. Sorry to bother you: Designing bots for effective recommendations. In *2019 IEEE/ACM 1st International Workshop on Bots in Software Engineering (BotSE)*, pages 54–58. IEEE, May 2019.
- [7] C. Brown and C. Parnin. Sorry to bother you again: Developer recommendation choice architectures for designing effective bots. In *2020 IEEE/ACM 2nd International Workshop on Bots in Software Engineering (BotSE)*. IEEE, July 2020.
- [8] D. Distefano, M. Fähndrich, F. Logozzo, and P. W. O'Hearn. Scaling static analyses at facebook. *Commun. ACM*, 62(8):62–70, July 2019.
- [9] S. H. Edwards and M. A. Perez-Quinones. Web-cat: Automatically grading programming assignments. *SIGCSE Bull.*, 40(3):328, June 2008.
- [10] S. H. Edwards, J. Snyder, M. A. Pérez-Quinones, A. Allevato, D. Kim, and B. Tretola. Comparing effective and ineffective behaviors of student programmers. In *Proceedings of the Fifth International Workshop on Computing Education Research Workshop, ICER '09*, page 3–14, New York, NY, USA, 2009. Association for Computing Machinery.
- [11] J. García, A. de Amescua, M. Velasco, and A. Sanz. Ten factors that impede improvement of verification and validation processes in software intensive organizations. *Software Process: Improvement and Practice*, 13(4):335–343, 2008.
- [12] N. Gitinabard, Y. Xu, S. Heckman, T. Barnes, and C. F. Lynch. How widely can prediction models be generalized? performance prediction in blended courses. *IEEE Transactions on Learning Technologies*, 12(2):184–197, 2019.
- [13] A. S. Hanks, D. R. Just, L. E. Smith, and B. Wansink. Healthy convenience: nudging students toward healthier choices in the lunchroom. *Journal of Public Health*, 34(3):370–376, 01 2012.
- [14] S. Heckman and J. King. Developing software engineering skills using real tools for automated grading. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education, SIGCSE '18*, page 794–799, New York, NY, USA, 2018. Association for Computing Machinery.
- [15] S. S. Heckman. An empirical study of in-class laboratories on student learning of linear data structures. In *Proceedings of the Eleventh Annual International Conference on International Computing Education Research, ICER '15*, page 217–225, New York, NY, USA, 2015. Association for Computing Machinery.
- [16] Z. Hu and E. Gehring. Use bots to improve github pull-request feedback. pages 1262–1263, 02 2019.
- [17] S. Kaleeswaran, A. Santhiar, A. Kanade, and S. Gulwani. Semi-supervised verified feedback generation. In *Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering, FSE 2016*, page 739–750, New York, NY, USA, 2016. Association for Computing Machinery.
- [18] D. G. Kay. Large introductory computer science classes: Strategies for effective course management. In *Proceedings of the Twenty-Ninth SIGCSE Technical Symposium on Computer Science Education, SIGCSE '98*, page 131–134, New York, NY, USA, 1998. Association for Computing Machinery.
- [19] A. M. Kazerouni, S. H. Edwards, T. S. Hall, and C. A. Shaffer. Deventracker: Tracking development events to assess incremental development and procrastination. In *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education, ITiCSE '17*, page 104–109, New York, NY, USA, 2017. Association for Computing Machinery.
- [20] J. C. Munson and S. G. Elbaum. Code churn: a measure for estimating the impact of code change. In *Proceedings. International Conference on Software Maintenance (Cat. No. 98CB36272)*, pages 24–31, 1998.
- [21] F. C. Payton and A. Busch. Examining undergraduate computer science participation in north carolina. *Commun. ACM*, 63(8):60–68, July 2020.
- [22] R. Rughiniş and S. Matei. Digital badges: Signposts and claims of achievement. In C. Stephanidis, editor, *HCI International 2013 - Posters' Extended Abstracts*, pages 84–88, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.
- [23] J.-G. Schneider, P. W. Eklund, K. Lee, F. Chen, A. Cain, and M. Abdelrazek. Adopting industry agile practices in large-scale capstone education. In *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: Software Engineering Education and Training, ICSE-SEET '20*, page 119–129, New York, NY, USA, 2020. Association for Computing Machinery.
- [24] L. Singer and K. Schneider. It was a bit of a race: Gamification of version control. In *2012 Second International Workshop on Games and Software Engineering: Realizing User Engagement with Game Engineering Techniques (GAS)*, pages 5–8, 2012.
- [25] G. Sprint and J. Conci. Mining github classroom commit behavior in elective and introductory computer science courses. *The Journal of Computing Sciences in Colleges*, page 76, 2019.
- [26] R. H. Thaler and C. R. Sunstein. *Nudge: Improving decisions about health, wealth, and happiness*. Penguin, New York, NY, USA, 2009.
- [27] R. H. Thaler, C. R. Sunstein, and J. P. Balz. *The behavioral foundations of public policy*, chapter Choice architecture, pages 428–439. Princeton University Press, Princeton, NJ, 2013.
- [28] D. R. Wallace and R. U. Fujii. Software verification and validation: an overview. *IEEE Software*, 6(3):10–17, 1989.
- [29] M. Weinmann, C. Schneider, and J. vom Brocke. Digital nudging. *Business & Information Systems Engineering*, 58(6):433–436, 2016.
- [30] C. Wilcox. The role of automation in undergraduate computer science education. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education, SIGCSE '15*, page 90–95, New York, NY, USA, 2015. Association for Computing Machinery.
- [31] C. Wilcox. Testing strategies for the automated grading of student programs. In *Proceedings of the 47th ACM Technical Symposium on Computer Science Education, SIGCSE '16*, page 437–442, New York, NY, USA, 2016. Association for Computing Machinery.
- [32] L. A. Williams and R. R. Kessler. The effects of “pair-pressure” and “pair-learning” on software engineering education. In *Thirteenth Conference on Software Engineering Education and Training*, pages 59–65. IEEE, 2000.