Student Paths in CS1: Case Studies of Initial Poor Performers

by

Ji Yeon Kim

Department of Computer Science
Duke University

Date: ______________________________

Approved:

____
Ashwin Machanavajjhala, Chair

____
Kristin Stephens-Martinez, Supervisor

____
Jeffrey Forbes

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science
in the Department of Computer Science
in the Graduate School of
Duke University

2019

ABSTRACT

Student Paths in CS1: Case Studies of Initial Poor Performers

by

Ji Yeon Kim

Department of Computer Science
Duke University

Date: ____________________________
Approved:

____
Ashwin Machanavajjhala, Chair

____
Kristin Stephens-Martinez, Supervisor

____
Jeffrey Forbes

An abstract of a thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science
in the Department of Computer Science
in the Graduate School of
Duke University

2019

Abstract

With the high influx of computer science enrollment in universities in the last decade, there is increasing value and wide-reaching effects in improving pedagogy in the field. This improvement is especially useful in introductory computer science courses (CS1). Student experience in the first programming course is known to heavily influence students' desires to stay in the field.

We present a set of student case studies that were enrolled in COMPSCI 101, an introductory computer science course at Duke University in fall of 2018. These case studies consist of students all with a poor initial performance in the course. We then consider their overall grade and help-seeking behavior in terms of what kind of help was sought and how often they sought help throughout their entire time in the course. We found no consistencies across any of the case studies.

Acknowledgements

I acknowledge and thank my advisor and mentor Kristin Stephens-Martinez, committee members Jeff Forbes and Ashwin Machanavajjhala, Susan H. Rodger for her insightful feedback, and program coordinator Marilyn Butler.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

It is important to understand how and when students in introductory computer science courses seek help. Such information will inform educators of effective study techniques, which will further improve computer science pedagogy. The high influx of computer science enrollment in universities in the past decade makes this contribution increasingly valuable. The Computing Research Association reported that in 2017, the number of students in the CS major in doctoral-granting universities in North America has tripled since 2006 [Ass17]. The steady increase in CS enrollment in four-year institutions means that contributions in improvements in CS pedagogy can have wide-reaching positive effects. In addition, previous research has shown that experience in the first programming course heavily influences students' desires to stay in the field, especially for female and minority students [BMK09, ST08].

A growing number of computer science courses leverage online course tools to help facilitate delivery of evaluations and assistance to students. These tools also allow instructors to access usage data of students. We analyzed student data from three course tools to predict final course grades and identify differences in help-seeking behavior among students. We specifically identified initially struggling students in order to conduct a case study analysis on those who did and did not improve their performance by the end of the course.

To select students for our case study analysis, we calculated their mid-semester grade with a linear regression based on gradebook information before or on the midterm. We found eight students with poor initial performance and conducted a case study with each student considering their performance on different parts of the course material, how much help they sought after the first midterm, and where they sought out this help. Two students were combined to result in seven case studies with no consistent behavior between any case study.

Chapter 2

Background and Related Work

Notable previous work in predicting student success in introductory programming courses include the Error Quotient [Jad06], Watwin score [WLG13], and Normalized Programming State Model (NPSM) [CHA15]. These models were able to achieve the highest accuracy in predicting student success with data obtained from student programs. For instance, the NPSM predicted student performance in an introductory programming course by analyzing student programming behavior, obtained through integrated development environment (IDE) logs. In contrast, work done by Bergin and Reilly [BR06] used a logistic regression with features such as student programming background, academic background, and mathematics background as opposed to student programming characteristics.

Studies conducted in this area repeatedly confirm that the most important predictors of success were student comfort level, self-efficacy, and mathematics background. [LYGE16, Wil02, Jr.05, WS01]. Lishinski et al. observed that self-efficacy impacts CS student performance [LYGE16]. Wilson identified effort invested by the student and comfort levels, among prior programming experience, gender, and other variables, to be the strongest predictors in success in an objects-first programming course [Wil02]. More studies concluded that student comfort level had the highest positive influence in success in an introductory programming course [Jr.05, WS01].

Our work is different from previous work in its objectives. The linear regression model was solely used to identify students that were predicted to be poor performers by the end of the course, where as previous work focused on creating accurate models to predict whether a student will succeed and identifying the highest predictors of success. Our work aims to observe similarities and differences in the way the initial poor performers, identified by a prediction model, seek help.

There has been previous work that analyzed help-seeking posts in an online discussion forum of a CS2 course [VBB+17]. It found that there was a correlation between overall grades and the category of the post. While our work does not show statistical significance in the relationship between the types of online discussion forum posts and overall grade, we base our overview of student paths on additional help-seeking channels, such as office hours and private tutoring attendance.

Chapter 3

Data Collection

We collected data spanning one semester (15 weeks) in fall of 2018 Duke University's COMPSCI 101, which is an introductory course on computer science in the Python programming language. The class had 281 students and we received consent from 171 students (60.9%) to use their data for research. We collected data from from three different sources. These sources were Piazza, a web discussion forum for students and teaching staff [Pia], an office hour queue app, and the gradebook.

3.1 Piazza

Piazza is a discussion forum where students enrolled in a course can ask questions or address concerns and elicit answers from teaching staff. As one of its features, Piazza provides us with downloadable Comma Separated Values (CSV) file comprised of the following:

- Student metadata (name, email, student ID)
- Post creation date (timestamp)
- Content of post (body of student's post)

3.2 Office hours app

The class used a web mobile app that managed an online queue for the students during office hours [SBF+17, CS6]. The class's office hours are run by undergraduate teaching assistants (TAs) and are a place where students can drop in and ask questions regarding course material. We collected the following information from the office hours app:

- Student metadata (name, email, student ID)
- Time student asks for assistance (timestamp)
- Summary of student's question or concern (title of student's post)
- Content of post (body of student's post)
- Steps student has taken to solve mentioned problem (body of student's post)

3.3 Gradebook

We collected the following information from the gradebook:

- Student metadata (name, email, student ID)
- Scores of every assignment for each student
- Scores of every textbook reading quiz for each student
- Scores of every algorithmic programming test (APT) quiz for each student
- Scores of every exam for each student

Chapter 4

Procedure

4.1 Selecting students

To choose which students to consider for our case study analysis, we calculated their mid-semester overall grades based on the weights of the syllabus (see Table 1). We used all gradebook information available up to the first midterm (the first six weeks of the course). We then used a linear regression to confirm how well this calculation matched the end of semester overall course grade. In our linear regression, we normalized each gradebook category. We did not use information from students who withdrew in our analysis to improve the model's accuracy. Training the regression with withdrawal students would introduce artificially low scores, because the missing scores from withdrawal students must be converted to scores of zero in the preprocessing step. We performed a 10-fold cross validation to prevent overfitting the model.

The coefficients for the linear regression are in Table 4.1. We found that the linear regression performed better than the syllabus weights and therefore used the results from the regression to choose our case study students.

4.2 Identifying Poor Performers

Among the 171 students, we identified eight students that were predicted to be a poor performers. Among these eight students, five did succeed, meaning that the

Table 4.1: The weight of each gradebook category from the course syllabus and coefficients from the linear regression model.

Graded category	Syllabus weight	Coefficient
Assignment total	20%	48.27
APT quiz total	9%	23.33
Peer instruction total	3%	3.68
Reading quiz total	3%	4.89
First midterm score	15%	38.81

Table 4.2: Table summarizing the number of students that were predicted to succeed versus predicted to not succeed and their actual outcomes.

	Actually succeed	Did not actually succeed
Predicted to succeed	157	6
Predicted to not succeed	5	3

students actually passed with a B- or higher. The remaining three did not succeed. Table 4.2 summarizes the students that were predicted to succeed versus not, and their actual outcomes.

Chapter 5

Case Studies

The following case studies focus on students predicted to perform poorly. We focused on eight students and created a total of seven case studies.

In order to situate students based on help-seeking behavior, we divided the levels of activity into three categories: no help-seeking, medium help-seeking, and high help-seeking. We defined these categories taking into account Smith et al.'s work on My Digital Hand [SBF+17], an online app for tracking office hour interactions. They found that the top 5% of students who most heavily used office hours used 50% of the total service time provided by the undergraduate TAs. We used this statistic to define high help-seeking as posting on Piazza or having a number of office hours sessions that fall within the top 5% of the class. If the number of posts from both channels were under the top 5% of the class, but greater than zero, the help-seeking activity was defined as medium. Finally, if there is no record of the student having attended office hours and no posts on Piazza, this was categorized as no help-seeking. The top 5% of students that most frequently used office hours had over 23 interactions with TAs and made over 19 questions to Piazza in the course of the semester. Figures 5.1 and 5.2 are empirical PDFs of the number of office hour sessions and number of Piazza posts, respectively. The activity levels and the respective number of events for each resource are provided in Table 5.1.

Among students that demonstrated some help-seeking, we additionally categorized the type of help-seeking with a question classification scheme created by Vellukunnel et al. [VBB+17], based on seminal work in learning categorization done by Chi [THC09]. The three types of questions we found were *active*, *constructive*, and *content-clarification*. *Active* questions request for help and omits information on reasoning as well as what the student has already tried. *Constructive* questions, on the other hand, "reflect students' reasoning or attempts to construct a solution to the problem" [VBB+17], as defined by Vellukunel et al. *Content-clarification* questions are requests for additional information on assignments, quizzes, and are not related to concepts in the course.

An overview of all case study students can be found in Table 5.2.

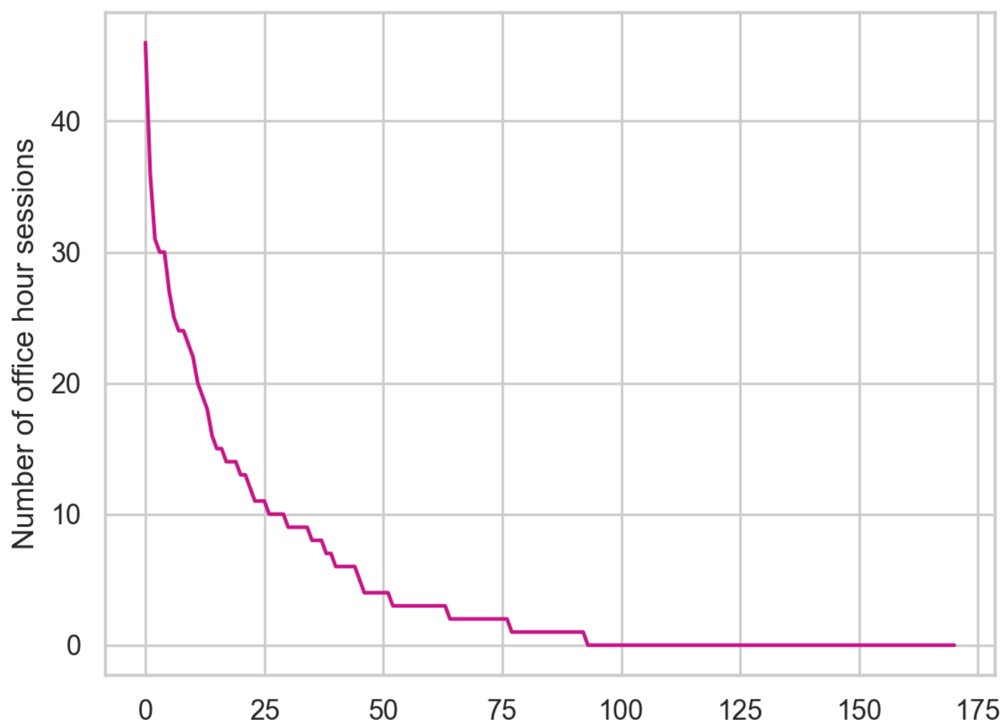

Figure 5.1: Empirical probabilistic density function of the number of interactions with TAs in office hours per student

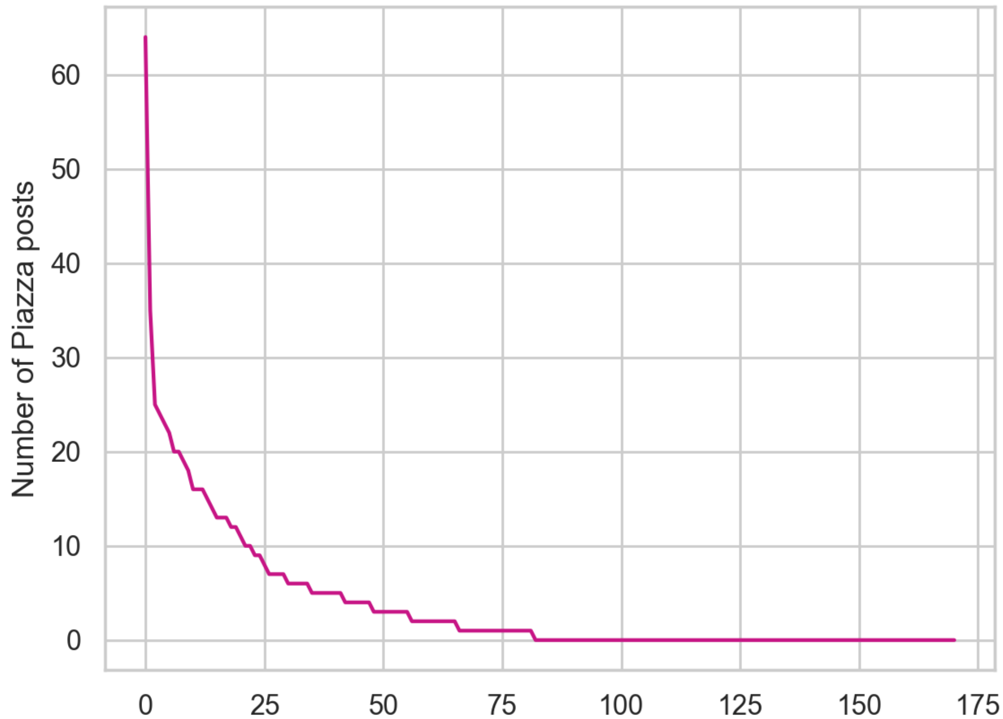

Figure 5.2: Empirical probabilistic density function of the number of Piazza posts per student

5.1 Case Study 1: No office hour activity, no Piazza activity, attended tutoring

Student A was predicted to score an overall grade of 78.9% by the end of the course. While the regression predicted they would perform poorly in the class, they were able to raise their overall grade to a 90.5 (A-).

Table 5.1: Percentage of students with N posts per help resource.

Activity level	Office hour sessions	Piazza posts
None	N = 0	N = 0
Medium	$0 < N \leq 22$	$0 < N \leq 18$
High	$N > 23$	$N > 19$

Table 5.2: Summary of case study students. Values in Predicted and Actual are predicted and actual overall grades of each student in percentages.

Case	Student	Office hours	Piazza	Tutoring	Pass	Predicted (%)	Actual (%)
1	A	None	None	✓	✓	78.9	90.5
2	B	None	Medium		✓	63.7	81.9
2	C	None	Medium		✓	78.6	86.9
3	D	Medium	Medium		✓	77.3	85.3
4	E	Medium	Medium			78.9	72.4
5	F	High	None	✓		70.2	61.2
6	G	High	Medium	✓	✓	78.4	87.3
7	H	High	Medium	✓		73.7	61.2

The first midterm exam score for Student A was 59.8%, which was in the 47th percentile among students that consented. This student had high assignment scores throughout the course except for the seventh assignment out of eight total assignments, which informs us that the first midterm exam score played the largest role in the low projected grade. The second midterm score improved to a 83.9%, which was in the 52nd percentile. On the final exam coding portion the student scored roughly above the mean with a 91.5%.

The interesting aspect of this student is the upward trajectory of course performance with absence of help-seeking from office hours and Piazza. This student was however was provided intervention after the first midterm with a targeted invite for a small private tutoring group. A group no larger than six students was given extra office hours with an peer instructor. Student A was reported to have filled out the form for private tutoring, and was subsequently added to a group which they attended for the remainder of the semester on a weekly basis.

5.2 Case Study 2: No office hour activity, medium Piazza activity, did not attend tutoring

Student B and C were both predicted by the regression to perform poorly by the end of the course, with projected grades of 63.7% and 78.6%, respectively. Student B raised the final course grade to 81.9 (B) and student C raised the final course grade to an 86.9 (B+). Student B barely made the cutoff for success while student C was moderately successful.

Both students had relatively high assignment scores throughout the course. The feature that had the highest factor in the predictions of the students' performances was the first midterm exam scores of 58.7% and 53.3% for student B and student C, respectively. These scores are within the bottom 4th and 5th percentile of first midterm exam scores among our students. Student B continued to maintain an upward trajectory in exam performance (69% for the second midterm, 81.9% for the final exam) as did student C (36.3% for the second midterm, 63.2% for the final exam).

These students are unique in our case study because they are the only ones to seek help exclusively on Piazza and not attend office hours. Student B posted a total of six times to Piazza and student C posted ten times throughout the semester. Student B asked questions regarding APTs and assignments and student C mainly asked about assignment autograder's output. Student B's questions were all active, and student C asked active, constructive, and content-clarification questions. One of student C's Piazza posts was a response to a question posted by a classmate. Both students were provided intervention after the first midterm with a targeted invite for a small private tutoring group. Both did not indicate interest in joining tutoring groups despite the targeted invitations. They are examples of students in our case study that were able to gain mastery of course material while having sought help only through Piazza.

5.3 Case Study 3: Medium office hour activity, medium Piazza activity, did not attend tutoring

Student D was predicted by the regression to score an overall grade of 77.3% in the course. Student D improved the trajectory of their performance and passed the course with an overall grade of 85.3% (B).

Student D had relatively high assignment scores throughout the course, above 50th percentile of all student total assignment scores. Again, the graded item that most heavily influenced the projected overall grade was the first midterm exam score of 54.3%, which was below the 15th percentile. There was an improvement in each subsequent exam score. Student D scored a 72.6% in the second midterm and 82.1% for the final exam.

This student is the only student in our case study set that had medium help-seeking activity in office hours and Piazza and received a passing overall grade. Student D posted once on Piazza and asked a total of six questions during office hours as indicated by the office hours online app. The subject of the questions asked during office hours ranged from exam preparation, to completing a function for an assignment, to understanding error messages. The questions were all active questions, because the student's response to a part of the form asking what they have tried were vague answers such as "having thought through the problem". In Piazza, the student asked a constructive question which revealed their understanding of what the return values of a function in an assignment should be. The student was sent a targeted invite for a small private tutoring group and was assigned a group, but ultimately did not attend any sessions. This student sought a medium amount of help through both office hours and Piazza and gained mastery of CS1 concepts. A potential alternative reason for their improvement may be that they sought help outside of class which we do not have records of.

5.4 Case Study 4: Medium office hour activity, medium Piazza activity, did not attend tutoring

In contrast to case study 3, student E was unable to improve their overall performance despite exhibiting very similar help-seeking behavior. Their predicted grade was 78.9% but their actual overall grade was 72.4% (C).

Student E had a below average cumulative assignment score up until mid-semester, just below the 40th percentile. The first midterm exam scores was 53.3%, which is within the bottom 10th percentile among first midterm exam scores of students that consented. These two items contributed to the relatively accurate overall grade prediction. Student E received a score of 36.3% in the second midterm and 63.2% for the final exam. While they did improve in exam performance after the second exam, it was not enough to demonstrate mastery of the course

Data shows that student E had medium office hour and Piazza activity. They asked a total of two questions during office hours and had a total of three Piazza posts, all after the first midterm. Student E asked strictly active questions. Posts to both help-providing channels solicited assistance in assignment completion rather than in grasping a concept. Student E did not indicate interest in joining a private tutoring group. This student is an example of one that did seek help in office hours, did not seek tutoring, and did not succeed in the class.

5.5 Case Study 5: High office hour activity, no Piazza activity, attended tutoring

Student F was predicted by the regression to perform poorly by the end of the course and indeed did not pass the course. The predicted grade was a 70.2% and the actual overall grade was a 61.2% (D-). Again, the feature that had the highest influence in the prediction of the student's overall performance was the first midterm exam scores of 29.3%, which is the lowest score among those of students that consented. This student is unique in our case study due to their exclusive use of office hours. Student F asked a total of 25 questions in office hours, all after the first midterm. All questions during office hours asked for help with completion of APTs or assignments and are characteristically active. Student F was sent a targeted invite for a small private tutoring group and was assigned a group. They attended at least one tutoring session. They are an example of a student that did not gain mastery of the course and attended at least one tutoring session and office hours.

5.6 Case Study 6: High office hour activity, medium Piazza activity, attended tutoring

Student G was predicted to score 78.4% in the course overall. However, Student G improved the trajectory of their performance and passed the course with an overall grade of 87.3 (B).

Student G exhibited a consistently upward trajectory in exam scores. For the first midterm, the student scored under the 5th percentile among consenting students, which was sufficient to significantly lower their projected overall score. They received as their second midterm score a 64.3% and 95.3% for the final exam.

Student G posted once to Piazza and asked a total of sixteen questions during office hours. This makes student G a medium Piazza user and high office hours user. The subject of the majority of the questions asked during office hours were regarding assignments and APTs. While most students forgo providing in the self-reporting form a detailed documentation of steps attempted to solve the problem they are asking, student G consistently provided this information throughout the semester. Student G's questions were therefore mainly constructive, with a few content-clarification questions. The student was also sent a targeted invite for a small private tutoring group and indicated interest in joining. They attended these sessions for the remainder of the course. Student G is an example that successfully improved their trajectory in the course whose records show they had high attendance in office hours, medium activity on Piazza, and attending tutoring sessions.

5.7 Case Study 7: High office hour activity, medium Piazza activity, attended tutoring

In contrast to student G in the previous case study, student H was unable to succeed in the course with similar levels of help-seeking behavior. Student H was predicted to score a 73.7% in the overall course grade but the actual grade was a 61.2%.

Average assignment score from student H was 89.9%, a score falling below the 10th percentile. The student also received a first midterm exam score of 33.7%. These two graded items significantly lowered the overall predicted grade. After receiving a score of 28.6% in the second midterm exam, student H received the lowest final exam score among those that consented.

Student H exhibited a high amount of help-seeking by asking a total of 24 questions during office hours and medium amount of help-seeking in Piazza with a total of five posts. All office hour and Piazza activity began after the first midterm. The subjects of the questions asked during office hours were completing or debugging assignments and APTs. The Piazza posts were comprised of questions regarding assignment autograder output. Student H asked characteristically active questions in both help-seeking channels. This student was also provided a targeted invite for a small private tutoring group and had attended the session at least once. This student is a counterexample from that of case study 6; student H ultimately could not improve their trajectory in the course, while records show they had high attendance in office hours, medium activity on Piazza, and attending tutoring sessions.

Chapter 6

Conclusion

We present seven case studies on eight students on their help-seeking behavior. These eight students all performed poorly in the first half of the course. We found no consistencies in the help-seeking behaviors across any of the case study students with regards to frequency and preferred help-seeking channel. Our case study results do show a positive correlation between students whose questions are constructive and their course performance, complementing results in a previous study by Vellukunnel et al. [VBB+17].

One limitation to this study is self-selection bias. Out of 281 students, only 60.9% of the students consented to release their data to be used for research purposes. While some explicitly rejected giving data, many did not even fill out the consent form. We suspect this reflects a significant statistical change in our data from the original dataset of 271 students.

In addition to the selection bias, our records for the number of Piazza posts and office hour sessions are not complete, and this is because we could not always link students' Piazza accounts to the rest of their data. Also, students were reported to sometimes not use the online app during office hours. Tutoring group records were also limited. There were tutoring groups organized outside of the teaching staff's knowledge, and we did not have record of who attended these campus-organized study groups.

The small data size does not allow us to make any statistically significant inferences. We also note that we should take into consideration confounders such as student health and course load in the same semester to make causal inferences.

Our data collection and procedure can be improved by incentivizing more students to participate in the study, urging students to use the online app when attending office hours, documenting the types of questions asked during private tutoring groups, and collecting data on the kinds of help students receive out of class.

One avenue for future work is studying students that show partiality for a specific channel for help. In our case studies we had students that used to office hours and went to Piazza. Students that actively go to office hours are less likely to succeed than those that exclusively use Piazza in moderation, or use Piazza in conjunction with office hours (case studies 5 and 7). Whether one channel promotes self-efficacy, one of the highest factors of student success in CS1, more than others can be an idea to be explored as future work.

Bibliography

[Ass17] Computing Research Association. Cra releases report on surge in computer science enrollments, February 2017.

[BMK09] Lecia J. Barker, Charlie McDowell, and Kimberly Kalahar. Exploring factors that influence computer science introductory course students to persist in the major. In *Proceedings of the 40th ACM Technical Symposium on Computer Science Education*, SIGCSE '09, pages 153–157, New York, NY, USA, 2009. ACM.

[BR06] Susan Bergin and Ronan Reilly. Predicting introductory programming performance: A multi-institutional multivariate study. *Computer Science Education*, 16:303–323, 12 2006.

[CHA15] Adam S. Carter, Christopher D. Hundhausen, and Olusola Adesope. The normalized programming state model: Predicting student performance in computing courses based on programming behavior. In *Proceedings of the Eleventh Annual International Conference on International Computing Education Research*, ICER '15, pages 141–150, New York, NY, USA, 2015. ACM.

[CS6] Cs 61a queue, April.

[Jad06] Matthew Jadud. Methods and tools for exploring novice compilation behaviour. *ICER 2006 - Proceedings of the 2nd International Computing Education Research Workshop*, 2006, 01 2006.

[Jr.05] Philip R. Ventura Jr. Identifying predictors of success for an objects-first cs1. *Computer Science Education*, 15(3):223–243, 2005.

[LYGE16] Alex Lishinski, Aman Yadav, Jon Good, and Richard Enbody. Learning to program: Gender differences and interactive effects of students' motivation, goals, and self-efficacy on performance. In *Proceedings of the 2016 ACM Conference on International Computing Education Research*, ICER '16, pages 211–220, New York, NY, USA, 2016. ACM.

[Pia] Piazza, April.

[SBF+17] Aaron J. Smith, Kristy Elizabeth Boyer, Jeffrey Forbes, Sarah Heckman, and Ketan Mayer-Patel. My digital hand: A tool for scaling up one-to-one peer teaching in support of computer science learning. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*, SIGCSE '17, pages 549–554, New York, NY, USA, 2017. ACM.

[ST08] Robert Sloan and Patrick Troy. Cs 0.5: A better approach to introductory computer science for majors. pages 271–275, 01 2008.

[THC09] Michelene T. H. Chi. Activeconstructiveinteractive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1:73 – 105, 01 2009.

[VBB+17] Mickey Vellukunnel, Philip Buffum, Kristy Elizabeth Boyer, Jeffrey Forbes, Sarah Heckman, and Ketan Mayer-Patel. Deconstructing the discussion forum: Student questions and computer science learning. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*, SIGCSE '17, pages 603–608, New York, NY, USA, 2017. ACM.

[Wil02] Brenda Cantwell Wilson. A study of factors promoting success in computer science including gender differences. *Computer Science Education*, 12(1-2):141–164, 2002.

[WLG13] C. Watson, F. W. B. Li, and J. L. Godwin. Predicting performance in an introductory programming course by logging and analyzing student programming behavior. In *2013 IEEE 13th International Conference on Advanced Learning Technologies*, pages 319–323, July 2013.

[WS01] Brenda Cantwell Wilson and Sharon Shrock. Contributing to success in an introductory computer science course: A study of twelve factors. *SIGCSE Bull.*, 33(1):184–188, February 2001.