# The Dangers of Participation Bias in Computer Science Educational Studies

Ryan Bockmon
The Roux Institute at Northeastern University
Portland, Maine, USA
r.bockmon@northeastern.edu

Aditya Jain
The University of Nebraska - Lincoln
Lincoln, Nebraska, USA
ajain9@unl.edu

Stephen Cooper
The University of Nebraska - Lincoln
Lincoln, Nebraska, USA
scooper22@unl.edu

## ABSTRACT

This paper investigates the extent to which participation bias impacts computer science educational studies. To explore the impact of participation bias, we ran two studies, the first study we had students volunteer to participate in filling out a pre-survey and the second study we made mandatory for all students as part of their course grade. Our results showed that when running a voluntary study, around 50% of the eligible student population chose to participate and those students that chose not to participate in this study were more than three times more likely to fail or withdraw from their course. In an effort to mitigate participate bias we conducted a follow up study where participation was mandatory for all students. Our results from this study showed that when making participation mandatory, the percentage of students who participated increased. However, 20% of students still chose not to participate even when participation was mandatory. Moreover, those students who did not participate are significantly more likely to fail or withdraw from the course. These results indicate that the majority of the at-risk student population are not participating in our studies and are not fully represented. More effort needs to be made in reducing participation bias to help capture at-risk students needs, as these students are the ones who are needing the most help but are not fully being represented in our studies.

## CCS CONCEPTS

• **Social and Professional topics → Student assessment**.

## KEYWORDS

CS1, Surveys, Participation

**ACM Reference Format:**
Ryan Bockmon, Aditya Jain, and Stephen Cooper. 2018. The Dangers of Participation Bias in Computer Science Educational Studies. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 7 pages. https://doi.org/XXXXXXX.XXXXXXX


ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

## 1 INTRODUCTION

Student success in introductory computing courses continues to be a major challenge. Failure and dropout rates have been reported to be as high as 40% in some cases [16]. Like many others, we wanted to study reasons why students would be at risk of failing or dropping out of an introductory computing course. Over several years of conducting these studies assessing students' attitudes and aptitudes in computing, we observed that approximately half of the targeted student population willingly took part in our studies for the pre-assessment, and among those, half participated in the post-assessment. The number of participants in these studies left us concerned about any results that we found. We suspected that students who volunteered to participate were students who were more likely to pass the course, more likely to have better attitudes towards computing, and stronger computing-related skills and that students who did not participate are more likely to fail or withdraw from the course. The exact student that we wanted to study.

In this paper, we explore two research questions:

(1) What is the extent of participation bias in a voluntary computer science educational study?
(2) To what extent can we mitigate participation bias in computer science educational studies by making participation mandatory?

## 2 RELATED RESEARCH

### 2.1 At-Risk Students in CS

Over the years, studies have examined what factors impact student success in computing courses and if it would be possible to predict students based on these factors before they failed or withdrew from the course [6, 7, 11–13, 17–19]. Researchers in these studies found that factors corresponding to demographics, attitudes, and aptitudes all could influence a student's ability to pass a computing course.

Werth found that, overall, college grades, the number of hours worked, and the number of high school math classes taken had the most significant relationship to course grades [17]. Wilson et al. found that students' comfort level and math grades had a positive influence on their success, while attribution to luck had a negative influence on their success [19]. Rountree et al. found that the only indicator of success from their surveys was whether or not a student was expecting to get an A from the course [13]. Byrne et al. found a clear link between programming ability and existing aptitude in mathematics and science subjects [7]. Wiedenbeck et al. found that both students' mental models of programming and self-efficacy have a direct effect on their overall success in an introductory programming course [18]. Lacher et al. ran a study across six sections of CS1 in the fall of 2015 and found a correlation between students aptitude level and final course grade [11]. Bergin and Reilly ran a study from 2003-2004 and found that students', gender, comfort level, mathematics scores, science scores and perception of their understating of a computational module accounted for 79% of the variance in programming performance [3].

A few studies have also looked into building models to predict at-risk students. In 2019 Liao et al. found that prerequisites and clicker responses provide high accuracy for predicting students' risk levels [12]. They also found that assignments and online quizzes added some accuracy to these predictions. When accounting for clicker questions, online quizzes and assignments, their model was able to achieve up to 79% accuracy when predicting students' risk levels, where risk levels is the opposite of student success. In 2020, Bockmon et al. found that students' spatial skills had the highest prediction of success in a computing course. Bockmon et al. also found that accounting for students' prior programming abilities, prior spatial skills, socioeconomic status, and several attitudinal factors such as, personal interest and fixed mindset resulted in a model that had an accuracy of 64.6% at predicting students' risk levels [6].

While the results found in these studies are significant based on the data that was collected and highlight important insights into what impacts students success, the results are biased based on students who participated filling out the surveys and does not capture students who did not.

## 2.2 Biases in Human-Based Research

The understanding and prevention of biases in human-based research is not new. It is common practice in most science fields to randomize experiments and participants to reduce selection bias [20]. When randomization is not feasible, researchers have been known to use propensity scores to help reduce selection bias [10]. In the medical field, it is common practice to run a double-blind study by having both the researcher and test subject not know if they are in the control or treatment group to help reduce the Hawthorne effect [14]. Studies have also investigated selection bias based on the tools and methods to collect data, such as having access to a computer or the Internet to fill out online surveys. Couper studied the issues and approaches of running web-based surveys and stated that running web-based surveys leads to selection biases where data is only collected from subjects that have access and knowledge to use a computer [8].

When it comes to human subject research, it is challenging to choose who decides to participate in these studies. Researchers can try to reduce selection bias by randomly picking which subjects partake in what study, but are generally not able to force subjects to participate. Even when participation is mandatory, subjects still have the right to choose not to participate, and few studies have looked into the impact of participation bias on educational studies.

Slonim et al. hypothesized people who participate in an economic lab experiment are not representative of the population [15]. In their study, they targeted 892 students who were enrolled in a first-year undergraduate introductory microeconomics class to participate in various economic lab experiments. They found that on average, there were 23% of targeted students participating in their experiments and that their participants were not representative of the target population on almost all characteristics, including having lower income, working fewer hours, volunteering more often, and exhibiting behaviors correlated with interest in experiments and economics.

Duquin et al. studied the potential influence of recruitment in neuropsychological research by re-analyzing data that they collected from an older study [9]. In that study they collected data across three different groups: paid research volunteers, clinical patients undergoing evaluation, and clinical patients referred for diagnostic clarification. They found that the research volunteers were significantly younger and less frequently cognitively impaired compared to the clinical groups, greatly affecting their estimates.

# 3 METHODS

## 3.1 Survey Administration

Two different studies were administered at a large R1 institution in the Midwest. The studies were aimed at undergraduate college students who were enrolled in an introductory computer science course. Two surveys were administered as a pre-assessment during the first two weeks of the semester. These two surveys were designed to collect students' programming readiness for introductory computer science and an attitudes survey to gauge students' attitudes towards computing-related topics [4, 5].

The first study was administered during the spring semester of 2021 and the second study was administered in the fall semester of 2021. The goal of the first study was to investigate the extent to which participation bias impacted the population of our study by looking at final letter grades between students who participated and those who did not participate in our study when participation was voluntary [2].

Student grades in the U.S. are protected under the Family Educational Rights and Privacy Act (FERPA), a federal law that protects the privacy of student education records. However, Section 99.31(6)(i) approves of disclosing student data if the "disclosure [of students' grades without consent] is to organizations conducting studies for, or on behalf of, educational agencies or institutions to (A) Develop, validate, or administer predictive tests; (B) Administer student aid programs; or (C) Improve instruction." [1]. Because we were in the process of developing and administering a predictive test, our university's IRB allowed us to access grades from all students in the course, not just from students who signed the study's consent form.

The first study was done by conducting voluntary pre-surveys that were administered during the first two weeks of the course to gauge students' prior programming abilities. Which has been common practice in our previous studies. The pre-surveys were administered in three different course sections. Students were incentivized to participate by receiving a small portion of extra credit towards their final course grade. There was a total of 459 students who were enrolled in one of the department's introductory computing courses that were eligible to participate in the study. Of these 459 students, 202 students signed up and completed the assessment instruments that made up our study. All students' final letter grades were obtained after the semester was over. Final letter grades were then compared between students who voluntarily participated in the study and those who did not. The use of final letter grades for both participating and non-participating students was approved by the university's IRB and the department chair.

The second study collected data during the fall semester of 2021. The goal of this study was to investigate making participation mandatory for all students in a study. The study was set up the same way as the first study. Pre-surveys were administered during the first two weeks of the course to gauge students' prior programming abilities. In this case, students were tasked with completing the survey as part of their course grade, either as a homework assignment or lab grade. They were not graded on how well they did on the survey, only graded on whether they completed the survey or not. There was a total of 563 that were enrolled and were eligible to participate in the study. Despite linking the pre-survey with a portion of the student's grade in the course, of these 563 students, 458 students completed the survey. Again, final letter grades were obtained for all students, those who completed the survey and those who did not.

## 3.2 Analysis

Two groups were tracked: the students who completed the pre-survey referred to as participants, and those who did not complete the pre-survey, the non-participants. The final letter grades for each student were converted to number-based grades, corresponding with the grade point average (GPA) scale defined by the office of the university's registrar. Students who withdrew from or did not complete the course were marked with a 0. No data was collected on why students might have withdrawn from the course. Withdrawal from the course could depend on various factors, such as personal issues, familial matters, financial stress, etc.

Measures of central tendency (means and medians) were computed for participants and non-participants from both studies. Bar charts and box plots were used to visualize the distribution of final letter grades for participants and non-participants. To determine any statistically significant difference, a T-test was used to compare their sample means if the data were normally distributed. If the data was not normally distributed, a Mann-Whitney U test was used, since it is a non-parametric alternative to the T-test. Throughout this study, an alpha of 0.01 is used as the cutoff for significance, and a p-value of less than 0.01 was used to indicate a statistical significance in the results.

# 4 RESULTS

## 4.1 Voluntary Study

In the first study, where participation was voluntary, there were a total of 202 participants and 257 non-participants. Tests for normality were run on the participants and non-participants to determine if the data set was normally distributed. Both the participants and non-participants were found to have a non-normal distribution for the two groups. A Mann-Whitney U was used to test for significant between the two groups. Results found that there was a significant difference between final letter grades of participants and non-participants (U-statistic = 33522 and a p-value of $< 0.01$). Table 1 and Figure 1 highlights the distribution of grades between participants and non-participants. The mean GPA for the participants was around a 3.15 and the median GPA for the participants was 4.0.

**Table 1: GPA (Mean and Medians) and U-Statistic between Participants and Non-Participants in a Voluntary Study**

| Participants Mean (Median) | Non-Participants Mean (Median) | U-Statistic (P-value) |
|---|---|---|
| 3.15 (4.0) | 2.4 (3.0) | 33522 (<0.01) |

While the mean GPA for the non-participants was around 2.4. The median GPA for the non-participants was 3.0.

**Figure 1: Box Plots of GPA Distribution of Participants and Non-Participants in a Voluntary Study**

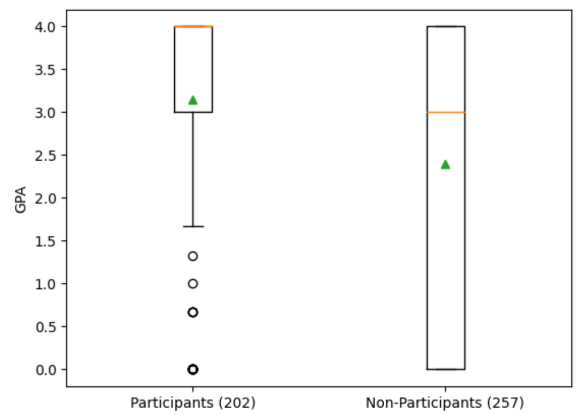
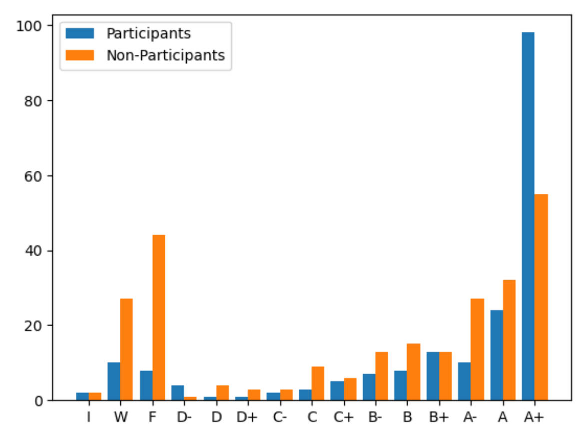



**Figure 2: Grade Distribution of Participants and Non-Participants in a Voluntary Study**




Illustrated in the bar graphs (Figures 2), there are significantly more participants receiving an A+ (98 students) compared to non-participants (55 students). Almost two times more students who participated in the survey received an A+ than those who did not participate. On the lower half of the grading scale, a significantly higher number of non-participants than participants earning letter grades at or below a C. Moreover, there are disproportionately more non-participants failing or withdrawing (73 students) compared to

**Table 2: GPA (Mean and Medians) and U-Statistic between Participants and Non-Participants in a Mandatory Study**

| Participants Mean (Median) | Non-Participants Mean (Median) | U-Statistic (P-value) |
|---|---|---|
| 2.88 (4.0) | 2.23 (2.67) | 30656.5 (<0.01) |

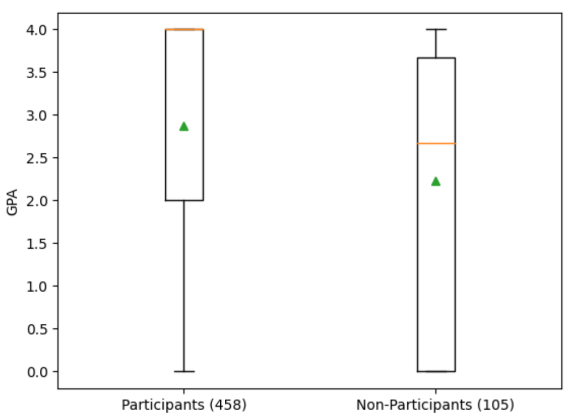
participants (20 students). Students who did not participate were more than 3.5 times more likely to fail or withdraw from the course.

## 4.2 Mandatory Study

In the second study, participation was mandatory for all students who were enrolled in an introductory CS course. There was a total of 563 students who were enrolled and a total of 458 students participated in the study and the remainder of the 105 did not participate. Even when making participation mandatory, there were still a significant amount of the students who chose not to complete the pre-survey (around 20%).

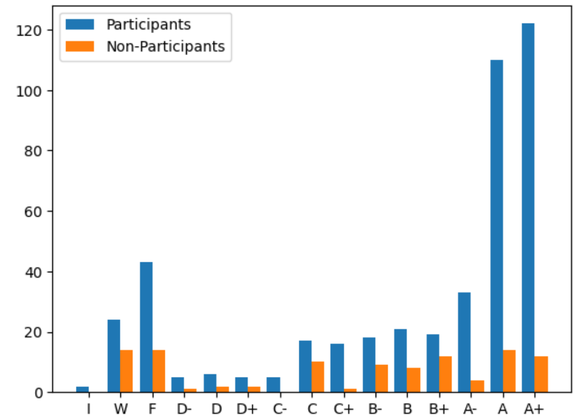
Both the participants and non-participants were found to have a non-normal distribution for the two groups. A Mann-Whitney U was used to test for significant between the two groups. Results showed that there was a significant difference between final letter grades of participants and non-participants (U-statistic = 30656.5 and a p-value of < 0.01). Table 2 and Figure 4 highlight the grade distribution. The mean GPA for the participants was 2.88 and the median GPA was 4.0. The mean GPA for the non-participants was 2.23 and the median GPA for the non-participants was 2.67.

**Figure 3: GPA Distribution of Participants and Non-Participants in a Mandatory Study**




The grade distribution between participants and non-participants is seen in Figure 4, and it highlights the difference in the number of students between the groups. The grade distribution follows a similar distribution pattern as the voluntary study (Figure 2). Where the participants are more skewed towards getting an A in the course. However, for non-participants, their distribution is mostly spread evenly across all grades. When comparing participants and non-participants, there is still a significant bias of participants receiving an A and non-participants withdrawing or failing. 235 out of 458

**Figure 4: Grade Distribution of Participants and Non-Participants in a Mandatory Study**




(51%) participants received an A or higher in the course, while 29 out of 105 (28%) non-participants received an A or higher. 30 out of the 105 (29%) of non-participants failed or withdrew from the course, while and 65 out of the 485 (14%) of participants failed or withdrew from the course. Even when participation was mandatory, about 20% of students chose not to participate, and those students who did not participate were more likely to fail or withdraw from the course.

## 5 DISCUSSION

The results of our two studies provide evidence for the extent of participation bias's impact on the data collected. In the voluntary study, administered in the spring of 2021, the effects of participation bias are significant as we observed a statistically significant difference in final letter grades between participants and non-participants. The participants in the voluntary study had statistically higher mean and median values for their GPA compared to those of the non-participants. Moreover, the participants' GPAs were overall concentrated in the upper half of GPAs as seen in Figure 1, while the non-participants' GPAs were spread throughout the entire possible range. This illustrates that the participants outperformed the non-participants, indicating that the sample from this study was not representative of our desired population.

In the second study, carried out in the fall of 2021, we hoped to diminish the effect of participation bias by making participation mandatory for all students. Our results suggest that while participation bias was indeed mitigated, it was not eliminated. Even when making participation mandatory, we had about 20% of students choosing not to participate, and these students who did not participate were statistically more likely to fail or drop out of the course.

A distinctive characteristic of our second experiment is that even with an effort to make participation mandatory, 105 students did not complete the pre-survey, choosing not to participate. This is a sizable portion of the total students, which sheds light on students' reluctance to participate, and the method used in each course section to try to obtain complete participation.

Participation bias is a prevalent issue in research studies that rely on optional human participation for data collection. It refers to the systematic differences between participants and non-participants in a study, which can lead to a biased sample that does not accurately represent the population of interest, as seen in our voluntary, first study despite its significance, participation bias is often mentioned briefly in the limitations section of research papers and not explored in depth. This practice can have serious implications for the generalizability of the study's findings that could lead to erroneous conclusions and recommendations. Therefore, in this paper, we argue that participation bias is more than just a formality that needs to be mentioned but must be acknowledged more seriously. Regardless of strategy, participation was still reliant on the student, and in our case, a large share of students chose not to participate. Through our efforts to mandate student participation, we did notice mitigated effects of participation bias, however, we were not able to eliminate it entirely.

## 6 LIMITATIONS

While our study addresses participation bias, it's essential to acknowledge potential confounding variables that could impact our results. Participants' motivation and incentives to participate may introduce biases beyond what we've directly measured. For instance, students who voluntarily engage in surveys might differ from non-participants in terms of their intrinsic interest in the subject matter or their desire for extra credit. These unmeasured factors could influence both participation and final outcomes, potentially affecting the observed relationship between participation and course performance.

Our methodology for carrying out our second study was imperfect because of the sizeable number of non-participants we had. A limitation related to our sample is the generalizability of our findings. Since our sample and population consisted of undergraduate college students in an introductory computer science course at a single university, generalizing our results too broadly would be imprecise and unsound. However, there is evidence to show our findings apply to our target demographic. There may also have been some confounding variables that could have led to inconsistency of overall student performance between the two semesters in which our two experiments were conducted.

Another limitation to our study was not being able to study the demographic of students who participated versus those who did not participate. We would have liked to understand the demographics of the students who did not participate in the study, but by the nature of this study they are not the ones filling out the surveys, and it would be unethical to gather their demographic data without their permission.

## 7 CONCLUSION

Our research aimed to identify the extent to which participation bias affects data in computer science educational studies; moreover, if there are strategic ways to reduce its impact. Over the course of two studies, with different methodologies, we were able to witness not only substantial effects of participation bias but also were able to help mitigate this participation bias by making participation mandatory. However, we were not successful in fully eliminating participation bias.

In the future, our study provides opportunities to investigate additional biases that could impact educational research outcomes. Two biases we are interested in exploring are motivation bias and incentive bias. Motivation bias may arise from students' varying levels of intrinsic motivation to participate actively in studies. Some students may engage enthusiastically, while others might lack interest or perceive participation as a mere obligation, tying into participation bias directly. Understanding how motivation impacts study outcomes can enhance the validity of findings. Similarly, incentive bias pertains to the effects of extrinsic rewards of participation, such as extra credit or assignment credit. While incentives encourage engagement, they may inadvertently attract specific student profiles, potentially skewing results. Our future research aims to explore these biases and their implications for study design, participant recruitment, and results interpretation.

Participation bias can manifest in any experiment that relies on human participation, and thus our findings concern a great deal of publications. Participation bias should no longer be concealed or overlooked. Rather, it must be acknowledged as it can have a grand impact on the results of a study, as it did in ours. Participation is a staple in countless publications' Methods section, and ultimately, participation bias is inherently found in their Limitations section. We urge readers, researchers, and academics alike that it is time to see eye-to-eye with participation bias in computer science, and consequently, strive to tackle it. Further work must be done to determine different methods to tackle and mitigate participation bias, whether it be through studying incentives, motivations, demographics, or through a mandatory strategy, as we took. With a developed understanding of this phenomenon and an active effort to minimize it, we will have more generalizable results, representative data, and improved findings.

## ACKNOWLEDGMENTS

We would like to thank XXXXXXX, for helping us collect data from his course.

## REFERENCES

[1] [n. d.]. Family Educational Rights and Privacy Act (FERPA). https://studentprivacy.ed.gov/faq/what-ferpa. Accessed: Aug 1, 2023.

[2] Anon Author 1 and Anon Author 2. [n. d.]. Anon Title.

[3] S. Bergin and R. Reilly. 2005. Programming: Factors that Influence Success. *SIGCSE'05* (2005), 411–415.

[4] Ryan Bockmon and Chris Bourke. 2023. Validation of the Placement Skill Inventory: A CS0/CS1 Placement Exam. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1* (Toronto ON, Canada) *(SIGCSE 2023)*. Association for Computing Machinery, New York, NY, USA, 39–45. https://doi.org/10.1145/3545945.3569738

[5] Ryan Bockmon, Stephen Cooper, Jonathan Gratch, and Mohsen Dorodchi. 2020. Validating a CS Attitudes Instrument. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education* (Portland, OR, USA) *(SIGCSE '20)*. Association for Computing Machinery, New York, NY, USA, 899–904. https://doi.org/10.1145/3328778.3366830

[6] Ryan Bockmon, Stephen Cooper, Jonathan Gratch, Jian Zhang, and Mohsen Dorodchi. 2020. Can Students' Spatial Skills Predict Their Programming Abilities?. In *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education* (Trondheim, Norway) *(ITiCSE '20)*. Association for Computing Machinery, New York, NY, USA, 446–451. https://doi.org/10.1145/3341525.3387380

[7] P. Byrne and G. Lyons. 2001. The Effect of Student Attributes on Success in Programming. *ACM SIGCSE Bulletin* 33, 3 (2001), 49–52.

[8] Mick P Couper. 2000. Web surveys: A review of issues and approaches. *The Public Opinion Quarterly* 64, 4 (2000), 464–494.
[9] J Aubrey Duquin, Brett A Parmenter, and Ralph HB Benedict. 2008. Influence of recruitment and participation bias in neuropsychological research among MS patients. *Journal of the International Neuropsychological Society* 14, 3 (2008), 494–498.
[10] Suzanne E Graham. 2010. Using propensity scores to reduce selection bias in mathematics education research. *Journal for Research in Mathematics Education* 41, 2 (2010), 147–168.
[11] L. Lacher, A. Jiang, Y. Zhang, and M. Lewis. 2017. Aptitude and previous experience in cs1 classes. *Int'l Conf. Frontiers in Education: CS and CE* 17 (2017), 87–91.
[12] S. N. Liao, D. Singaro, C. Alvarado, Griswold W. G, and L. Porter. 2019. Exploring the value of different data sources for predicting student performance in multiple CS courses. *SIGCSE' 19* (2019), 112–118.
[13] N. Rountree, J. Rountree, and A. Robins. 2002. Predictors of Success and Failure in a CS1 Course. *ACM SIGCSE Bulletin* 34, 4 (2002), 121–124.
[14] Philip Sedgwick and Nan Greenwood. 2015. Understanding the Hawthorne effect. *Bmj* 351 (2015).
[15] Robert Slonim, Carmen Wang, Ellen Garbarino, and Danielle Merrett. 2013. Opting-in: Participation bias in economic experiments. *Journal of Economic Behavior & Organization* 90 (2013), 43–70.
[16] Christopher Watson and Frederick WB Li. 2014. Failure rates in introductory programming revisited. In *Proceedings of the 2014 conference on Innovation & technology in computer science education*. 39–44.
[17] L. H. Werth. 1986. Predicting student performance ni a beginning computer science class. *Proceedings of the seventeenth SIGCSE technical symposium on Computer science education* (1986), 138–143.
[18] S. Wiedenbeck, B. LaBelle, and V. Kain. 2004. Factors affecting course outcomes in introductory programming. *Workshop of the Psychology of Programming Interest Group* (2004).
[19] B. C. Wilson and S. Shrock. 2001. Contributing to success in an introductory computer science course: a study of twelve factors. *SIGCSE 2001* (2001), 184–188.
[20] Christopher Winship and Robert D Mare. 1992. Models for sample selection bias. *Annual review of sociology* 18, 1 (1992), 327–350.