The Future of Engineering Education
2024 Annual Conference & Exposition
Oregon Convention Center
Portland, OR . June 23 - 26, 2024
ASEE
Paper ID #44308

Investigating How Student Attributes and Behaviors Relate to Learning Outcomes in a Free Online Python Programming Course

Mr. Timothy James, Purdue University

Timothy James is an instructor at the University of Pittsburgh Computer Science department, as well as a Doctor of Technology student at Purdue University. Previously, Tim has spent some time in a variety of industries including Internet consulting, finance, defense contracting, aviation maintenance, telecommunications, capital markets, and sandwich artistry. Timothy hopes to continue actively engaging the community in technical training and CS education efforts.

Dr. Alejandra J. Magana, Purdue University

Alejandra J. Magana, Ph.D., is the W.C. Furnas Professor in Enterprise Excellence in the Department of Computer and Information Technology and Professor of Engineering Education at Purdue University.

Investigating How Student Attributes and Behaviors Relate to Learning Outcomes in a Free Online Python Programming Course

Abstract

Students learning a programming language in a free, online environment are faced with several challenges - beyond the difficult material, the content must hold their attention and keep them coming back when there is no credit and there are minimal repercussions for failure or withdrawal. Attrition rates are high in these types of courses, and reducing attrition could have positive benefits. Determining student attributes and behaviors that could improve success may be valuable in helping many students learn a new programming language and could help to meet the high demand for computer science education.

The authors enrolled 921 students from around the world in a voluntary, noncredit, introductory Python programming course across several cohorts in 2022 and 2023. While these courses contained minor experimental variations for research purposes, the focus, topics, content, and evaluation criteria were similar. Student participation and completion were evaluated for each course. Surveys were administered to enrolled students that gathered data on experience, intent, behaviors, and demographics. Responses to these surveys indicate a racially diverse group of students with varying ages, levels of experience, educational backgrounds, and programming confidence.

This paper presents the student demographic data collected and aims to analyze these attributes to determine whether any of these factors correlate with higher rates of student success in these courses, measured by student participation rate and completion rate. Better understanding of these qualities may be used to encourage future cohorts of students and improve student achievement. This understanding may also be used to improve curriculum design so that future courses are able to effectively engage a broader audience.

1. Introduction

The number of jobs in software development is projected to increase substantially over the next decade [1]; this increased demand will require many new workers to learn how to develop software. Traditionally, many universities and colleges have provided computer science degree programs that will prepare future workers. However, more scalable approaches like Massive Open Online Courses (MOOCs) could be an alternative – a more scalable approach to preparing the next generation of software developers that might reach a broader audience [2]. These courses can help to address rising demand for computer programming education and expand access to educational opportunities [3]. Unfortunately, MOOCs suffer from high attrition rates [2] [4] [5]. If factors that improve the chances of student success in this type of course could be identified, they could be used to reduce attrition rates and improve educational outcomes in a more scalable fashion.

The purpose of this research is to understand if identified student attributes and behaviors are related to higher levels of success in a free, online, voluntary, noncredit, introductory Python programming course. The course was developed by the authors and provided to over 900 students in several cohorts, with the same general curriculum delivered online via Google Classroom over a period of 18 months. Students in these courses were evaluated using multiple-choice quizzes, participation in reflection exercises, programming assignments, and a final exam. Some of these students did not participate, some participated but not complete all requirements, and some successfully completed the requirements for the course. Before, during, and after these courses, students were asked questions about themselves and their experience. Student responses were reviewed for possible relationships between surveyed attributes and participation and completion rates.

2. Background

Both the salary and predicted demand for software developers is high [1]. Unfortunately, learning programming is not easy. Many computer science concepts are challenging, including programming; variables, program context, and logical pathways introduce a high level of cognitive load [6]. Difficult concepts make it hard for students to learn in any environment.

Online environments present their own challenges beyond the material and content presented. Unfamiliar tools and platforms can introduce additional difficulty to student. Further, the relationships that are often established in a traditional in-person course between students and teachers, as well as peer relationships, may not exist as creating these personal connections in online environments can be an additional challenge [7]. Students' continued engagement in these types of courses is noted to be a complex phenomenon with many factors [5]. These impediments, among others, likely contribute to high student attrition. This is often observed to be at least 90% [2] [4]; even at Georgia Tech in a paid, for-credit course, attrition was noted to be more than 23% [8].

Previous attempts have been made to identify relationships between student demographics and outcomes. Some of these studies use demographics to correlate with outcomes [9] [10] but do not focus on programming education. Other similar studies focus on programming education [11] but focus on how different demographics navigate these courses. Still others explore student motivations in MOOCs and how these relate to student demographics [12]. This analysis should add to the literature by focusing on programming education and relating both demographics and behaviors to student outcomes.

3. Definitions

For the purposes of this study, a student is considered to have **enrolled** if they accepted the invite to join the course in Google Classroom. Not all students who sign up accept the invitation to join the course in Google Classroom. A student is said to have **participated** in the course if they enrolled in the course and submitted any graded item: a single quiz, homework assignment, or reflection exercise. A student is said to have **completed** the course if they earned 75% or more of the points available. **Attrition** refers to any student who was enrolled in the course but does not complete the course.

4. Methods

13 introductory Python programming courses were provided between May 2022 and October 2023. Evaluation of students consisted of graded assignments, quizzes, reflection exercises, and an exam. While there was some limited variation in the content, or the manner in which that content was presented in these courses, they all covered approximately the same introductory material on Python programming. There were no restrictions on student sign-up; students could be any age, any level of experience, and in any location worldwide.

As part of the sign-up process, students were presented with a demographic survey that included the following questions:

- Expected Time Spent: How much time would you be able to spend engaging the material in this course?
- Programming Confidence: On a scale of 0 to 5, how confident are you in your programming skills?
- Programming Experience: How many years of experience do you currently have in computer programming?
- Age: How old will you be (in years) as of [the course start date]?
- Location: What is your current country of residence?
- Formal Training: Have you received formal training in any programming language from a bootcamp, college, university, or other higher education institution?
- Education Level: What is your highest completed level of education?
- Gender: What is your gender?
- Race / ethnicity: What race / ethnicity best describes you?
- Native English Speaker: Is English your native language?
- Employment Status: Which of the following best describes your employment status?

More than 90% of the students who signed up for the course provided responses to these questions.

In December 2023, surveys were sent to past enrollees in order to further investigate why some students do not complete the course and why others do not participate in the course. Survey questions to non-participating and non-completing students were similar, and asked students to rate their agreement and disagreement with statements on a 5 point Likert scale, with 1 being “Strongly disagree” and 5 being “Strongly agree”. These statements included:

- "I did not have enough time to fully participate in the course."
- "Technical challenges (unrelated to the course content) kept me from fully engaging the course."
- "I was not able to complete some course requirements or assignments because they were too confusing."
- "I was not able to complete some course requirements or assignments because they were too difficult."
- "Unexpected events in my life prevented me from completing the course."
- "Other stresses in my life prevented me from completing the course."
- "I did not feel that I could get help when I had questions."

20 (5.13%) of the students from the non-participating group and 50 (13.5%) students from the non-completing group responded to the survey.

4.1 Participants

The course was free to sign up; a form was shared publicly by posting an announcement to LinkedIn and Twitter for registering for an online course. 921 students who were enrolled across these 13 courses are the focus of this analysis. Of the 921 enrolled, 390 (42.3%) did not participate. 371 (40.3%) students participated but did not complete the courses. 160 students completed the courses (17.4%).

According to the demographic information provided, 521 (56.57%) of the participants identified as male, 330 (35.83%) of the participants identified as female, and 70 (7.6%) of the participants identified as non-binary or did not self-identify.

Self-identified race / ethnicity included 344 Black students (37.4%), 213 Asian students (23.1%), 136 Caucasian / White students (14.8%), and 80 Hispanic / Latino students (8.7%). There were not 20 or more students self-identifying as any other race / ethnicity.

A majority of students (502, or 54.5%) reported being based in the US. Several other countries were represented, including Ghana (107, or 11.6%), Nigeria (97, or 10.5%), and India (34, or 3.7%). There were not 20 or more students self-identifying from any other countries.

5. Results

All tests for statistical significance were performed using a two-tailed z-test.

In these courses, students self-identifying as female were significantly (95% confidence) more likely to participate (62.73% vs 55.47%), but gender did not seem to play a large role in completion rates. While a slightly larger percentage of students who self-identified as male completed the course (18.81% vs 16.67%), this is not a statistically significant result.

Native English speakers were more likely to participate in the course that non-native English speakers (60.69% vs 54.92%, respectively) and were more likely to complete the course (20.69% vs 14.39%). This difference in completion rates is statistically significant (95% confidence). As the course is authored in English by a native English speaker, this is not a surprising outcome, but care could be applied in future iterations to make the content more effective for a broader audience.

Native English Speaker	**Number of Students**	**Participated**	**%**	**Completed**	**%**
Yes	580	352	60.69%	120	20.69%
No	264	145	54.92%	38	14.39%
Unknown	77	34	44.16%	2	2.60%

Participation rates were similar for students in all reported employment statuses, but students who reported being unemployed completed the course at a higher rate than others (28.89%). When the 134 students who reported being unemployed are compared with the 831 who reported the other employment statuses (Full-Time, Part-Time, Student, or Other / unknown), the difference in completion rates – 28.89% for unemployed students, vs 16.13% for all others – is statistically significant (99% confidence).

Employment Status	**Number of Students**	**Participated**	**%**	**Completed**	**%**
Employed Full-Time (32 hours or more per week)	333	201	60.36%	70	21.02%
Unemployed	90	54	60.00%	26	28.89%
Employed Part-Time (less than 32 hours per week)	78	46	58.97%	7	8.97%
Student	329	182	55.32%	47	14.29%
Other / unknown	91	48	52.75%	10	10.99%

Students were asked to rate their programming confidence on a scale of 0 to 5, with 0 being “Not at all confident” and 5 being “Extremely confident” – as confidence increases, the completion rate increases. Perhaps unsurprisingly, student confidence is highly correlated with student completion rates (Pearson correlation coefficient = 0.934). Participation rates were also strongly correlated with self-reported student confidence (Pearson correlation coefficient = 0.674), but notably less so than completion rates.

Self-Rated Confidence	Number of Students	Participated	%	Completed	%
0	206	107	51.94%	20	9.71%
1	204	120	58.82%	36	17.65%
2	193	120	62.18%	35	18.13%
3	164	95	57.93%	38	23.17%
4	74	45	60.81%	18	24.32%
5	23	14	60.87%	9	39.13%
Unknown	57	30	52.63%	4	7.02%

As illustrated in the table below, prior training in programming or computer science appears to be related to both higher participation rates and higher completion rates. Participation rate and completion rate differences are statistically significant (99% confidence) when comparing students reporting having received prior training to those who reported not having received prior training.

Prior Training	Number of Students	Participated	%	Completed	%
Yes	426	274	64.32%	97	22.77%
No	423	217	51.30%	58	13.71%
Unknown	72	40	55.56%	5	6.94%

Programming experience, however, is less clear as indicator of student success than prior training. Students stating that they have some experience (2-5 years) participate and complete the course at higher rates than more experienced or less experienced groups. The completion rate of this group is statistically significant (99% confidence) when compared to the less experienced group (14.94% vs 31.78%). This may indicate that students early in their learning journey (but not *too* early) have a higher appetite for this type of course, but more research and analysis are necessary to confirm this.

Years of Programming Experience	Total	Participated	%	Completed	%
Less than 2	716	413	57.68%	107	14.94%
2 or more, but less than 5	129	75	58.14%	41	31.78%
5 or more	52	30	57.69%	11	21.15%
Unknown	24	13	54.17%	1	4.17%

Students who self-identified as having the least education (have not completed high school) and students who self-identified as having the highest level of education (doctoral degree) completed the course at higher rates than all other education levels. However, the numbers of students enrolled from these groups is low; these participation and completion rates are not statistically significant when compared with students that reported all other education levels, so it is imprudent to attempt to draw a deeper conclusion from these data points. Participation rates and completion rates from the other education levels (Masters Degree, Bachelors Degree, High School, and Associates Degree) were similar to each other.

Highest Education Level	**Total**	**Participated**	**%**	**Completed**	**%**
Doctoral Degree	24	16	66.67%	7	29.17%
Masters Degree	165	86	52.12%	31	18.79%
Bachelors Degree	435	252	57.93%	80	18.39%
Associates Degree	52	31	59.62%	7	13.46%
High School	190	115	60.53%	27	14.21%
Have not completed high school	24	15	62.50%	7	29.17%
Unknown	31	16	51.61%	1	3.23%

As with education level, students at the highest and lowest self-reported ages completed the course at the highest rates. Participation rates and completion rates were similar for other ages.

Age Range	**Total**	**Participated**	**%**	**Completed**	**%**
Under 18	25	16	64.00%	8	32.00%
18-21	179	111	62.01%	26	14.53%
22-25	234	136	58.12%	43	18.38%
26-30	225	126	56.00%	38	16.89%
31-35	131	71	54.20%	20	15.27%
36-40	46	22	47.83%	5	10.87%
Over 40	55	36	65.45%	19	34.55%
Unknown	26	13	50.00%	1	3.85%

The previous attributes discussed are ones that students have little immediate ability to affect, but the question: “For 4 weeks, how much time would you be able to spend engaging the material in this course?” does provide students with some control. Students who stated that they would spend 2 hours or more had the highest rate of participation and completed the course at the highest rate.

Expected Weekly Study Time	Total	Participated	%	Completed	%
120 minutes or more per week	445	263	59.10%	85	19.10%
60-119 minutes per week	272	150	55.15%	45	16.54%
30-59 minutes per week	166	98	59.04%	27	16.27%
Less than 30 minutes per week	14	7	50.00%	2	14.29%
Unknown	24	13	54.17%	1	4.17%

Note that this is not a measure of how much time students spend on the course material – it is merely each student's own assessment of how much they intend to spend on the course material. An interesting observation is that the more time students *intend* to spend on the course material, the more likely they are to complete the course successfully. This result is not statistically significant; nevertheless, when comparing the minimum number of minutes students indicated they would study (120, 60, 30, and 0) to completion rates (19.10%, 16.54%, 16.27%, and 14.29%), there is strong correlation (Pearson correlation coefficient = 0.979). Correlation was lower between minimum number of minutes and participation rate, but still strong (Pearson correlation coefficient = 0.68).

Follow-up surveys submitted by students who successfully completed the course reinforce this idea; 52.2% of students stated they spent more than 2 hours on the course content (with an additional 17.4% stating they spent 60-119 minutes, and the remaining 30.4% stating they spent at least 30 minutes).

Another factor that students can control is whether they modify code. The lessons, provided as interactive computational notebooks, allow students to read, execute, and modify code. Starting in January 2023, students were asked the following question as part of a reflection exercise in the first week of the course: "Did you modify the code in the examples this past week?"

Students who reported modifying the code in the examples completed the course at twice the rate of students who reported not modifying the code; this result is statistically significant (99% confidence).

Modified Code?	Total	Completed	%
No	84	16	19.05%
Yes	275	112	40.73%

As part of each lesson in this course, students are given ungraded exercises. Completion of these exercises is another factor that indicates higher completion rates. Starting in January 2023, students were asked the following question as part of a reflection exercise in the second week of the course: "How much do you agree with this statement: "I have been attempting to solve the Try It! exercises at the end of the lessons."" Students answered using a 1-5 Likert scale, with 1 being "Strongly Disagree" and 5 being "Strongly Agree". While most students responded with a

5 – indicating strong agreement – these students completed the course at almost twice the rate of students who may have not spent as much time on these. This result is statistically significant (99% confidence).

Solving Try It! Exercises	**Total**	**Completed**	**%**
1-4	66	22	33.33%
5	97	57	58.76%

5.1 Student Participation and Completion Outcomes

While this data indicates that prior formal training correlates with higher student participation, there were no other attributes that clearly illustrate increased participation.

However, several of the demographic attributes in this data correlate with higher completion rates. Some of these attributes that are outside of the students' own immediate control include native language, employment status, other formal training, and programming experience. Students that reported English as their native language, being unemployed, having received prior formal training, and with 2-5 years of programming experience completed at higher rates. Notably, 8 student participants met all of these criteria. 7 of those 8 participated in the course, and all 7 of those participating students completed the course.

In addition, there are other factors that are under student control. While differences in students' expected time spent on the course material did not demonstrate statistical significance, the time spent was highly correlated with completion rates. In addition, interacting with provided code and examples – both through modifying code in lessons as well as attempting to complete ungraded exercises – led to significantly higher completion rates.

5.2 Student Perceptions

Follow-up surveys completed by both non-participating students and non-completing students indicated the same major challenges in engaging the course.

- Not enough time.
    - 80% of the non-participating and 72% of the non-completing respondents agreed or strongly agreed that they did not have enough time to engage the course.
- Unexpected life events.
    - 70% of the non-participating and 72% of the non-completing respondents agreed or strongly agreed that unexpected life events affected their engagement.
- Stress.
    - 65% of the non-participating and 68% of the non-completing respondents agreed or strongly agreed that other stresses in their lives affected their engagement.

Other issues were seen to be less of a factor in both the non-participating and non-completing groups.

- Course difficulty.
  - 85% of the non-participating and 70% of the non-completing respondents disagreed or strongly disagreed that the course content was too difficult to engage.
- Assistance.
  - 80% of the non-participating and 72% of the non-completing respondents disagreed or strongly disagreed that they were unable to get help.
- Confusion.
  - 75% of the non-participating and 68% of the non-completing respondents disagreed or strongly disagreed that confusing course requirements prevented them from engaging the course.

Responses were mixed regarding technical challenges unrelated to the course. 45% of the non-participating students agreed or strongly agreed this limited their engagement; the remaining 55% disagreed or strongly disagreed. Similarly, 28% of the non-completing students agreed or strongly agreed this was an issue, and 64% disagreed or strongly disagreed.

6. Conclusion and Future Work

As stated in the introduction, attrition in MOOCs is typically high. In this free, voluntary, noncredit, introductory Python programming course delivered online, attrition is similar; more than 82% of the enrolled students did not complete the course requirements. However, several trends can be seen when reviewing student attributes and behaviors and comparing these to participation rates and completion rates. Some of these are outside of student control, including prior formal training, native language, programming experience, or employment status. While employment status may affect how much time students have available for study, care could be applied to future iterations of the course to improve its understandability – both for non-native English speakers as well as more novice programmers who lack prior formal training or previous programming experience.

Based on these findings, several interventions could be made with students, and these could be explored in future research:

- As non-native English speakers completed the course at lower rates, additional support provided by native speakers, or better translations into students' native languages, may affect outcomes.
- As most students who successfully completed the course reported spending at least 2 hours per week on the course material, encouraging students to spend at least this much time may improve students' chance for successful completion. This may be especially helpful with students enrolled in other programs, or students with demanding work schedules.

- Supporting documentation or resources could be suggested for students with no prior training or less experience to enable them to catch up with other more experienced students.
- Students who attempt to modify code in the lessons complete the course at higher rates, so providing several reminders to modify and experiment with the code throughout the course could improve student understanding and lead to more course completions.
- Students who attempt to solve ungraded practice problems complete the course at higher rates, so encouraging students to solve these types of practice problems may result in higher completion rates.

Student opinions did not indicate that course difficulty or confusion were major factors in not completing the course, but that lack of time to focus on the course was an impediment. This is reinforced by student responses regarding the amount of time they intend to spend on the course.

Additionally, it may be worth noting that since prior formal training and higher levels of student confidence are correlated with higher student completion rates, successfully completing a course such as these may prepare students for success in future attempts to learn programming, but investigating or confirming this is beyond the purpose of this analysis.

This work inspires several questions that could be interesting for future study. Further investigation and research could illustrate if interventions like those suggested above have the intended effects. Additionally, it may be worth exploring the causal relationship. It is possible that students who are dedicated to completing the course may be more likely to spend more time, modify code, and attempt the ungraded exercises; further work would be necessary to attempt to identify if especially motivated students perform these activities, or if performing these activities leads to higher performance for any student. Finally, investigating other similar courses and their outcomes and aggregating results may help to understand more general applicability of these factors.

References

[1] U.S. Bureau of Labor Statistics, "Software Developers, Quality Assurance Analysts, and Testers," https://www.bls.gov/ooh/computer-and-information-technology/software-developers.htm.

[2] G. Conole, "MOOCs as disruptive technologies: strategies for enhancing the learner experience and quality of MOOCs," *Revista de Educación a Distancia (RED)*, no. 50, Jul. 2016, doi: 10.6018/red/50/2.

[3] I. Roll, D. M. Russell, and D. Gašević, "Learning at Scale," *Int J Artif Intell Educ*, vol. 28, no. 4, pp. 471–477, Sep. 2018, doi: 10.1007/s40593-018-0170-7.

[4] H. Fake and N. Dabbagh, "Personalized Learning Within Online Workforce Learning Environments: Exploring Implementations, Obstacles, Opportunities, and Perspectives of Workforce Leaders," *Technology, Knowledge and Learning*, vol. 25, no. 4, pp. 789–809, Dec. 2020, doi: 10.1007/s10758-020-09441-x.

[5] J. Goopio and C. Cheung, "The MOOC dropout phenomenon and retention strategies," *Journal of Teaching in Travel and Tourism*, vol. 21, no. 2, pp. 177–197, 2021, doi: 10.1080/15313220.2020.1809050.

[6] B. B. Morrison, L. E. Margulieux, and M. Guzdial, "Subgoals, context, and worked examples in learning computing problem solving," in *ICER 2015 - Proceedings of the 2015 ACM Conference on International Computing Education Research*, Association for Computing Machinery, Inc, Jul. 2015, pp. 267–268. doi: 10.1145/2787622.2787733.

[7] M. Yarmand, J. Solyst, S. Klemmer, and N. Weibel, "It feels like i am talking into a void: Understanding interaction gaps in synchronous online classrooms," in *Conference on Human Factors in Computing Systems - Proceedings*, Association for Computing Machinery, May 2021. doi: 10.1145/3411764.3445240.

[8] D. A. Joyner and C. Isbell, "Master's at scale: Five years in a scalable online graduate degree," in *Proceedings of the 6th 2019 ACM Conference on Learning at Scale, L@S 2019*, Association for Computing Machinery, Inc, Jun. 2019. doi: 10.1145/3330430.3333630.

[9] N. P. Morris, S. Hotchkiss, and B. Swinnerton, "Can demographic information predict MOOC learner outcomes?" [Online]. Available: http://eprints.whiterose.ac.uk/86184/

[10] Q. Zhang, F. C. Bonafini, B. B. Lockee, K. W. Jablokow, and X. Hu, "Exploring Demographics and Students' Motivation as Predictors of Completion of a Massive Open Online Course," *International Review of Research in Open and Distance Learning*, vol. 20, no. 2, pp. 140–161, 2019, doi: 10.19173/irrodl.v20i2.3730.

[11] P. J. Guo and K. Reinecke, "Demographic differences in how students navigate through MOOCs," in *L@S 2014 - Proceedings of the 1st ACM Conference on Learning at Scale*, Association for Computing Machinery, 2014, pp. 21–30. doi: 10.1145/2556325.2566247.

[12] R. Y. Bayeck, "Exploratory study of MOOC learners' demographics and motivation: The case of students involved in groups," *Open Praxis*, vol. 8, no. 3, p. 223, Jul. 2016, doi: 10.5944/openpraxis.8.3.282.