Who Continues in a Series of Lifelong Learning Courses?

Sami Sarsa
Aalto University
Espoo, Finland
sami.sarsa@aalto.fi

Arto Hellas
Aalto University
Espoo, Finland
arto.hellas@aalto.fi

Juho Leinonen*
University of Helsinki
Helsinki, Finland
juho.leinonen@helsinki.fi

ABSTRACT

Although computing education research quite often targets within-university courses, an important role of universities is educating the public through open online lifelong learning offerings. Compared to within-university courses, in lifelong learning, the student population is often more diverse. For example, participants often have more varied motivations and aspirations as well as more varied educational backgrounds. In this work, we explore what kinds of learners attend open online lifelong learning programming courses and what characteristics of learners lead to completing courses and proceeding to subsequent courses. We examine student-related factors collected through surveys in our online course environment. These factors include motivation, previous experience, and demographics. Our results show that motivations, previous experience, and demographics by themselves only explain a small amount of the variance in completing courses or continuing to a subsequent course. At the same time, we identify individual factors that are more likely to lead to learners dropping out (or continuing) in the courses. Our study provides further evidence that lifelong learning benefits most the already educated part of the population with prior knowledge and high motivation. This calls for further studies that seek to identify means to engage and support participants less likely to continue in such courses.

CCS CONCEPTS

• **Applied computing** → **E-learning**; **Distance learning**; **Interactive learning environments**; *Computer-assisted instruction*; *Computer-managed instruction*; • **Social and professional topics** → **Computer science education**; *Adult education.*

KEYWORDS

lifelong learning, open online courses, retention, motivation

ACM Reference Format:
Sami Sarsa, Arto Hellas, and Juho Leinonen. 2022. Who Continues in a Series of Lifelong Learning Courses?. In *Proceedings of the 27th ACM Conference on Innovation and Technology in Computer Science Education Vol 1 (ITiCSE 2022), July 8–13, 2022, Dublin, Ireland*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3502718.3524752

*Also affiliated with Aalto University.

ITiCSE 2022, July 8–13, 2022, Dublin, Ireland

ACM ISBN 978-1-4503-9201-3/22/07…$15.00
https://doi.org/10.1145/3502718.3524752

1 INTRODUCTION

As society is getting more digitized, technological literacy such as understanding basic computing concepts has become more important. These skills, sometimes referred to as twenty-first century skills, have been argued to be essential to be able to participate e.g. in the modern workforce [20]. Universities have risen to the occasion by increasing the availability of their courses to the general public of lifelong learners. In courses that are specifically aimed for lifelong learners, such as massive open online courses (MOOCs), the student population is often different when compared to traditional university classrooms [4, 8, 15, 18], where most students are young adults in their early twenties.

In this work, we study the characteristics of attendees participating in a series of courses focusing on principles of computer science through programming aimed for lifelong learners in a Nordic country. We study the demographics of the participants (e.g. gender, age, prior educational qualifications), motivations for attending the courses, as well as previous experience in programming. Additionally, we explore the characteristics of students who continue to the end of the course, and of those who take a subsequent course in the course series. We study to what extent the characteristics could be used to predict who will continue in order to understand who are at the biggest risk of dropping out, and thus would potentially most benefit from e.g. adjusting the contents of the course series or changing the way the courses are conducted.

This work contributes to prior works that have studied participants' demographics and retention in online courses (e.g. [9, 22]). While previous studies have often focused on one or more individual courses instead of course series, we look also into characteristics of those lifelong learners that continue beyond a single course. Our overall research theme is seeking to understand *who joins and persists in a series of lifelong learning courses.* From the computing education research domain, close matches to our work are e.g. [5, 28, 29], although their emphasis is in MOOCs.

2 BACKGROUND

2.1 Lifelong Learning

Lifelong learning is not limited to classrooms or educational institutions, nor it is limited to specific age groups or learning after attending an educational institution [10]. The term lifelong learning has been around for a long while, and these days it is often also associated with governmental policies [11]. As an example of a governmental (and cross-governmental) policy, the year 1996 was proclaimed as the "European Year of Lifelong Learning" by the European Union, during which one of the missions was to promote lifelong learning [39]. At the core of these policies is the view that education is one of the drivers of economic growth and that education plays a role in sustainable and inclusive economic growth [40].

These views are dominant even though studying the phenomenon is challenging [1] and some have pointed out that higher productivity might be achieved faster by adapting practices from other countries [7].

These policies can, in part, influence education. While it is relatively common to have government-level incentive models such as providing funding to universities based on published research and completed degrees [12, 19], some countries also encourage universities to provide education to the general populace (e.g. Finland[1] where the study was conducted). This is also the case in the country where this study has been conducted: a part of the government-level funding is distributed across the universities depending on the number of credits attained by non-university students. In practice, this means that there are real incentives for trying to attract non-university students and for trying to increase course retention rates of those students.

2.2 MOOCs and Distance Education

One of the major trends from the last decade are MOOCs (massive open online courses) that at the peak of the hype were touted to revolutionize and disrupt education [31]. When compared to the long-standing tradition of distance education that has sought to bridge the physical distance between a student and a teacher [33], two of the key components of MOOCs are that they are offered free of charge and that they can scale to very large audiences through technical means. Realizing this was not trivial, however, and early studies on the topic discussed challenges with offering such courses [27]. Similarly, while MOOCs were originally thought to bring higher education to masses, research suggests that they primarily serve the already educated population [4, 8, 15] and that while the enrollment rates are high, the completion rates are often rather poor [32]. MOOCs have also been criticized for being one-shot courses that forget the broader picture [42].

From a broader perspective, MOOCs are a part of a larger phenomenon of building online learning offerings, which in turn are a natural evolution of distance education. Thus, it is perhaps not surprising that they suffer from similar challenges; as an example, completion rates have been a concern in both distance education [18] and online education [26, 37]. As Howell et al. [18] aptly voice out, these comparisons are not always straightforward or even sensible however, as – for example – the student populations are typically different to on-campus students [18]. Thus, one stream of research has sought to identify factors that contribute to student retention in online offerings.

2.3 Course Retention and Inclusion

A variety of factors are associated with course retention and academic performance [17]. When considering online education, in addition to course activity [3, 6], multiple factors including age, level of education, prior experience with the topic, relevance of the topic, intent to complete the topic and motivation have been identified to contribute to retention [13, 24]. When considering open CS courses and CS courses in general, one of the challenges is inclusion of traditionally underrepresented groups [41], including women who are often a minority in online CS courses [14, 22, 23, 25]. To combat this problem, researchers have proposed methods for increasing retention and for building more inclusive courses [2, 23], including reconsidering the difficulty of (early) course exams and improving course contents [2] and bringing best practices from locally offered CS courses online [23].

3 METHODOLOGY

3.1 Context and Data

This study was conducted in a series of open online lifelong learning courses organized by Aalto University. The course series focuses on principles of computer science and programming. Presently, there are four courses in the series, (1) introduction to programming, (2) data and information, (3) internet and browser applications, and (4) mobile application development. Each course is worth 2 ECTS, which corresponds to approximately 50 to 60 study hours and in total to approximately 200 to 240 study hours.

The courses are created for lifelong learning, and thus the contents, expectation level, and breadth somewhat differs from normal university courses – this has been done in the light of making the content more approachable to learners, which has been shown to reduce dropouts [2]. The used technologies are relatively new, which could allow those already familiar with programming to also learn something. The courses use Dart as the programming language and Flutter as the framework for building mobile applications. The courses are currently offered in Finnish. Programming exercises in the courses can be solved directly in the browser within the course environment, are automatically assessed, and participants can attend an online discussion forum for asking questions. The courses do not have explicit deadlines and participants can work on them at their own pace.

The courses are graded pass / fail, and passing a course means completing at least 90% of the available exercises. On course completion, participants are eligible to official university credits. Participants can also work on the courses anonymously without registration, but in this case, they cannot apply for credits.

As the learners work on the course contents, the platform periodically – based on progress – shows a survey that asks for some additional details about the learners. Answering the surveys is optional, and the questions are on motivations for attending the course series (n = 1004), demographic factors (n = 547), and prior programming experience (n = 392), where n in parentheses corresponds to the answers received to those questions. The questions are summarized in Table 1 in the results section.

3.2 Research Questions and Approach

For the present study, our research questions are as follows.

- RQ1 What characteristics (demographic, prior experience, motivations) do learners attending courses aimed at lifelong learning for programming possess?
- RQ2 How are these characteristics related to performance in an introductory programming course, and continuing to a subsequent course?
- RQ3 What is the relationship between gender, prior programming experience and completing the introductory and a continuation course?

[1] https://okm.fi/en/steering-financing-and-agreements

We answer the research questions through statistical analysis of the data. First, we show and analyze the descriptive statistics of learner retention in the courses in order to answer the first question.

For the second question, we analyze the effect of characteristics on predicting course outcomes using machine learning models. For this, we use regression with the sum of completion percentages for the introduction course and the furthest completed subsequent course for a given learner as the target[2], hereafter referred to as *total completion*. Models used for regression in this study comprise Linear regression, and Bayesian Ridge regression as linear models, Random forest as non-linear regression model, and a dummy regressor Mean regression (always predicts the mean of training data target) as a baseline to determine the effectiveness of our regression models. The assumptions for using regression are fulfilled (i.e. the variables are not highly correlated, apart from multicollinearity within groups due to Dummy Variable Trap).

In addition to predicting the continuous total completion with regression models, we use classification models to predict a) binary completion of the introductory course (90% of points required) and b) continuing to a subsequent course (10% of points in any subsequent course required). We use Logistic regression as a linear model for classification, and Random forest as a non-linear model for classification. As the classification baseline we have Majority vote, which always predicts the most common label and is equivalent to Mean regression for classification.

We use 5 fold cross-validation for both the regression and the classification tasks and report the variability of the resulting coefficients of the best performing models in each task in order to analyze feature importance.

For the third question, we take a closer look on how gender and previous experience affect course completion and continuation by visualizing the distinct group dropouts separately. We also compute test statistics with the non-parametric Kruskal-Wallis H test for previous experience differences separately for different genders and report ϵ_R^2 estimate of effect size [21, chapter 22] for the H statistic and also root mean-squared standardized effect size Ψ [36], which is analogous to the more commonly used Cohen's d, but for multiple distributions. We do not use parametric tests such as one-way ANOVA since the course completion percentage distribution is hardly normal.

3.3 Data Preprocessing

To form the dataset used for this study, we used the following preprocessing steps. First, we excluded underage learners due to privacy regulations. Second, we excluded unregistered learners. Third, since we are studying continuing from introductory course, we excluded learners who started working on another course prior to completing the introductory programming course. Finally, we remove all learners who had been active in the course platform during the last 30 days to exclude those who are still likely working on the course (in order to not bias the data). After these steps, we had 2801 learners in the data. Out of the 2801 learners, 392 have answered all the surveys[3] given by the platform.

[2]The sum has a possible overlap for learners who only completed introduction course with 90% of points and continued to a subsequent course.

[3]Roughly one third of those who were shown the last survey answered all the surveys.

4 RESULTS

4.1 Characteristics of Learners

To answer RQ1, we present descriptive statistics of those who participated in the studied courses. Table 1 shows the percentages of the learner population who gave a specific answer in the surveys. Looking at any single row in the table, one can see how the proportion of learners who answered a certain way changes over time – with the first value (30%) representing the proportion of learners who gave that answer when 30% of the introductory course was completed (the first point in time from which we have demographic data), the second (90%) representing the proportion of learners who gave that answer out of those who had completed at least 90% of the introductory course, and the third (10%) representing the proportion of learners who gave that answer out of those who continued to a subsequent course and completed at least 10% of the points there. For example, looking at the first row of the Table 1, 22.8% of the students at the 30% completion mark are 18 to 25 years old, while their proportion has decreased to 20.4% at the 90% completion mark. This indicates that the proportion of the 18- to 25-year-olds decrease over the course, meaning that they have a slightly higher dropout rate than the other age groups on average.

Based on Table 1, we see that younger learners are more likely to drop out as their proportion of the course population decreases. Similarly, we see that more experienced learners tend to be more likely to complete and continue as the proportions of those who have taken more courses, written more lines of code prior to the course, and/or whose self-estimated experience is higher increase over the course. We also see that more men start the course compared to other genders, and the gap increases when considering those who progress further (in the introductory course and to a subsequent course). Considering learners' educational background, we see that most learners who enroll in the courses have some tertiary education (63.4%), and that the proportion of those with some tertiary education increases when considering learners who progress further. Lastly, considering motivations, we see that for almost all the different motivations asked about in the survey, the proportion of learners who gave those specific motivations for attending the course decreases. Motivations are different from the other variables as learners could choose as many (or few) of the motivations as they wished. One possible interpretation here is that learners who continue further choose fewer motivations for attending the course compared to those who drop out early.

4.2 Learner Characteristics and Course Retention

To answer RQ2, we analyze the coefficients and effectiveness of the best models for predicting our three targets; total completion (introduction completion + a subsequent course completion), completing the introductory course, and continuing to a subsequent course.

For the first prediction task, we see from Table 2 that all machine learning models clearly best the Mean baseline, yet, the errors are large and explained variance quite small even for the best performing model Bayesian ridge regression. This suggests that the used

Table 1: Surveyed characteristics and their distributions for various learner cohorts. Lower index i stands for introduction course completion and s for subsequent course completion.

group	characteristic	$30\%_i$	$90\%_i$	$10\%_s$
age	18-25	.228	.204	.156
	26-35	.308	.291	.286
	36-55	.333	.363	.442
	56-	.031	.028	.030
	undisclosed	.099	.114	.085
courses taken	0-1	.524	.376	.296
	2-4	.306	.389	.421
	5-10	.170	.235	.283
gender	woman	.361	.284	.251
	man	.508	.574	.623
	other or undisclosed	.131	.142	.126
education	other inapplicable or undisclosed	.120	.139	.126
	secondary education or less	.245	.219	.207
	some tertiary education	.635	.642	.667
lines of code[1]	0	.373	.243	.208
	1	.416	.482	.465
	2-4	.211	.274	.327
motivation	it is free	.194	.173	.176
	course was recommended to me	.044	.021	.005
	for future career	.207	.159	.151
	interested in the topic	.453	.422	.467
	other or undisclosed	.066	.055	.055
	relevant to current role	.064	.042	.045
	to complete a university course	.130	.131	.111
	to learn about a specific technology	.272	.242	.281
self estimated[2]	1-2	.437	.310	.264
	3-4	.344	.394	.396
	5-9	.219	.296	.340

The number of learners in the groups vary. The proportions are calculated using all learners who answered the specific survey. The numbers of learners who answered the surveys are outlined at the end of Section 3.1.

[1] 0 = none, 1 = less than 500, 2 = 500-5000, 3 = 5001-40000, 4 = over 40000

[2] on scale of 1 (not at all experienced) to 9 (very experienced)

features are related to the course completion but not enough to accurately predict individual learner outcomes.

According to the Bayesian ridge regression coefficients shown on the left in Figure 1, the most distinctive feature group for predicting total completion is courses taken. Out of all the factors, courses taken 0-1 is the most negative factor by a large margin and the most positive factors are courses taken 2-4 and courses taken 5-10. While the effect is less prominent in other previous experience related variables groups and level of education, they all show the same pattern: the lower the experience, the smaller the coefficient.

Different motivation features also show some effect on the predicted value. Most notably, yet to no surprise, interest in the topic is the most positive motivation feature. The effect of external motivation features varies. To complete a university course is neutral while, for future career and relevant to current role is negative.

When inspecting how age or gender affects the total completion, we see that young or high age has a negative effect compare to ages

Table 2: Total completion regression scores

model	MAE	RMSE	explained variance	R^2
mean	0.53	0.60	0.00	-0.14
linear	0.46	0.55	0.11	0.02
bayesian ridge	**0.45**	**0.54**	**0.16**	**0.07**
random forest	0.46	0.57	0.05	-0.03

R^2 is calcutated using https://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2_score.html.

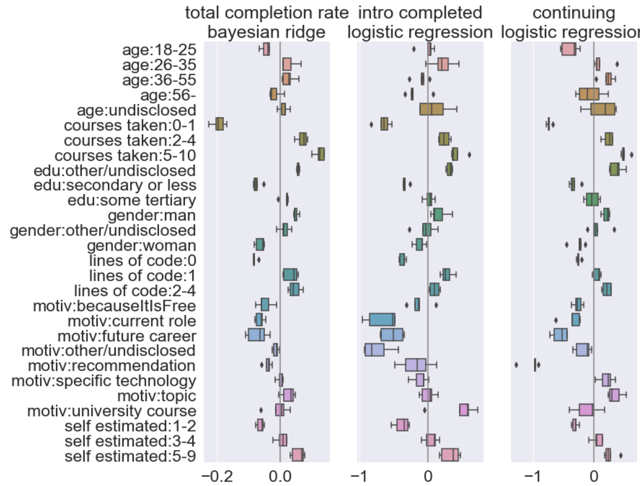

Figure 1: Variability of best model coefficients in 5 fold cross validation for prediction tasks

in between (26-55), and being a man has a positive effect as opposed to being a woman. This suggests that there is clearly work to be done in this aspect of the course series. For the other/undisclosed gender, the effect is neutral.

Similarly to the total completion prediction task, the best model for classifying learners into either completing or not completing the introductory course is a linear model, Logistic regression, although just barely, which can be seen on the left in Table 3. The machine learning models also outperform the majority vote baseline clearly in accuracy but receive only narrowly better RMSE scores and slightly worse F1-scores. Apart from F1-scores, the exact same can be said for the task of predicting continuing to a subsequent course.

The Logistic regression coefficients for the introductory course completion prediction task show largely a similar pattern as for predicting the total completion which can be seen in the boxplot in the middle of Figure 1. There are however, multiple evident differences. Most notably, courses taken 5-10 is not the greatest positive factor, but motivation for completing a university course, ages 36-55 appear negative for introduction completion but for total completion they seem a positive characteristic, and few lines of code experience is more positive indicator than extensive lines of code experience. Also, interest in the topic appears to be a neutral factor even though it is the second most positive feature out of the motivation features.

Table 3: Classification scores for predicting introduction completion (90% of points) and continuing (10% of points)

model	introduction				continuing			
	Acc	AUC	F1	RMSE	Acc	AUC	F1	RMSE
majority vote	0.58	0.50	**0.73**	0.50	0.59	0.50	0.00	0.50
logistic regression	**0.65**	**0.70**	0.70	**0.47**	**0.67**	**0.73**	**0.56**	**0.46**
random forest	0.62	0.65	0.70	0.50	0.63	0.68	0.52	0.48

Table 4: Kruskal-Wallis H test for previous experience explaining total completion percent for different gender groups

group	woman			man			other/undisclosed		
	p	ϵ_R^2	Ψ	p	ϵ_R^2	Ψ	p	ϵ_R^2	Ψ
courses taken	0.002	0.091	0.505	~0.0	0.220	0.618	0.041	0.277	0.686
lines of code	0.189	0.023	0.176	~0.0	0.150	0.535	0.057	0.250	0.775
self estimated	0.062	0.039	0.249	~0.0	0.107	0.441	0.188	0.145	0.567

Turning our attention to the right-most boxplot in Figure 1 depicting coefficients for predicting continuing to a subsequent course, we see that unlike in the first classification task, ages 36-55 is a positive factor. Further, being recommended the course is a much more negative factor and being interested in the topic is a more positive factor than in the other tasks.

4.3 Group Comparisons

To answer RQ3, we examined in more detail the effects and interplay of gender and prior experience on course completion. Table 4 shows Kruskal-Wallis H test p-values and effect sizes of the effect of previous experience on total completion percent separately for different gender groups. We see that for men, previous experience seems to better explain total completion percent: the p-values for all three previous experience variables are significant for men ($p < 0.05$), while only "courses taken" is significant for women and other/undisclosed. Similarly, the effect sizes are larger for men than women, while they are largest for other/undisclosed. These results suggests that prior experience is more strongly related to total completion percent for men compared to women. For other/undisclosed, prior experience is more strongly related (bigger effect size), but less statistically significant (higher p-values).

Figure 2 visualizes learner retention percentages, focusing on specific four background variables: gender, courses taken, lines of code and self-estimated prior experience. Looking at gender, we see that more women drop out compared to men. Other/undisclosed are between men and women. For the three experience related variables, we see that the dropout rate is the highest for those with less prior experience for all three variables.

5 DISCUSSION

When considering the course series overall, we observe a similar dropout phenomenon seen in other open online courses, despite the course series being targeted and built for lifelong learning. Considering views which posit that university-level introductory programming courses expect too much from their students [30], it is possible that the expectations in the course series should be

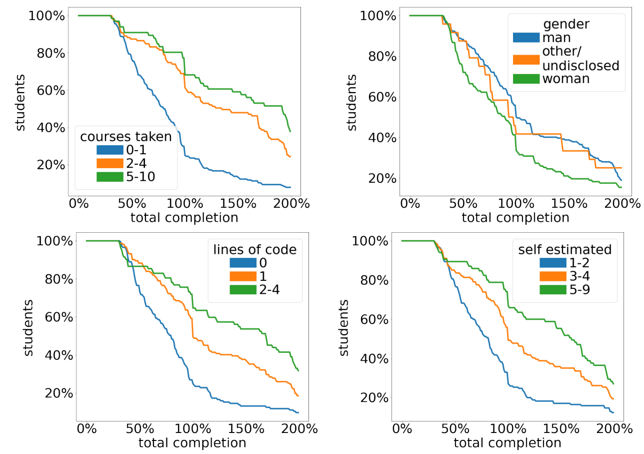

Figure 2: Learner retention percentages for different groups. The x-axis represents total course completion rate, i.e., sum of introductory programming course (up to 1.0) and any continuation course (up to 1.0), totalling together from 0.0 to 2.0.

decreased. At the same time, as it is possible that many of the learners are there to simply check the topic out, it is also possible that difficulty of the topic is not the reason why learners drop out, which leads to our analysis.

The most prominent results in our learner characteristic examination for retention is that those with previous experience are more likely to complete and continue in our lifelong learning course set. Even though most of our starting learners have little experience, this finding alongside that most of the participants already have some tertiary education, backs the notion raised by Reich et al. [32] that "MOOCs are primarily a complementary asset for learners within existing systems". Although lifelong learning MOOCs seem to suffer the same problems as MOOCs in general, the proportion of learners with secondary or less education could still be considered significant and the majority being already educated may not be an issue as one of the targets of lifelong learning is people with interest to re-educate themselves. After all, over third of our starting learners are already 36 or over and the mid-high age correlates with continuing to a subsequent course in our setting, indicating that such population appears to have been reached to reasonable extent. Furthermore, as computing education is again becoming popular in K-12 education in the Nordic countries (e.g. [16, 38]), the number of people with lack of experience in programming will likely decrease over the next decade.

Another notion is that the prediction scores of the models are lacking, which implies that motivation, previous experience and education are not enough to explain differences in completion rates. From the perspective of inclusivity, this is a good result; if gender and prior experience would be enough to explain variance in retention, it would suggest that the courses are very much tailored for people that have specific backgrounds.

Regarding the traditional split of intrinsic and extrinsic motivation, our results show somewhat mixed results. While our only clearly intrinsic motivational factor (interested in topic) overall appears to be a positive factor, the various external motivations show large differences. Especially wanting to complete a university course, which can be seen as a strong external motivation, is a much more positive factor for introductory course completion than intrinsic motivation[4] – we note however, that intrinsic motivation has not consistently been linked with programming course completion in university contexts either [34, 35]. On the other hand, future career appears to be the most negative motivational factor, which may have its roots in those wanting a future career in programming being less experienced and possibly having also otherwise worse starting conditions such as background in a less cognitively demanding profession or unrealistic expectations of programming. At the same time, wanting to attend the course due to learning for current role is also a negative factor, which is somewhat surprising. One possibility is that such participants are already proficient in the topic and choose to pick and mix content of interest to them from the courses, not completing all parts, which would show as them not completing the courses.

6 LIMITATIONS OF STUDY

Our study has multiple limitations. A major one is that due to avoiding survey fatigue, we ask for the demographic information only at the 30% mark. This means that our analysis of feature importance is limited to only learners who have completed the first 30% of the introductory course (and thus had the chance to answer all the surveys). This causes sampling bias as the populace of those who drop out earlier than that is likely different from the ones who reach the final survey threshold. Another sampling bias risk arises from the fact that filling the surveys is optional. Although, the exclusion of the first 30% of learners might not be an issue as the majority of those who drop out earlier might be just exploring. However, as we do not have the data, we cannot verify the differences between the dropouts within the first 30% of course points and those who continue further.

We also acknowledge that our definition of dropping out is a simple heuristic and some learners do continue work after a month's break. Also, content has been added for some of the courses gradually, which leads to a possibility that some learners may have dropped out (by our definition) in the middle of the course while some parts have not been published yet. Such learners would count as only having part of the full points. Further, the courses are first to offer learning materials for Dart and Flutter, which are emergent technologies, in the Finnish language. This may cause further bias in the population that attends the courses since it can attract people interested in these technologies specifically (approximately 28% of our continuing learners have marked this as one of their motivations). We have not gone deep into details in the characteristics of the learners and how these characteristics interact and affect each other. Even though we consider it interesting and beneficial, it is out of scope for this work. Rather, we provide a broad overview of the demographic, experience and motivational factors that contribute to continuing and completing a course.

[4]It is possible that some of the participants are students at universities that allow including credits from the courses to their degree.

7 CONCLUSION

In this article, we studied who continues in a series lifelong learning courses. Using data from a course series teaching principles of computer science, we outline three key observations. First, when considering the learners who complete at least 30% of the introductory programming course (which is the point where learners have been asked about their backgrounds), we find that the learners predominantly have little to no prior experience in programming, they already have attended tertiary education (or are currently attending tertiary education), and they are interested in the topic. Second, when considering who completes at least 90% of the introductory programming course (which is the point where learners can be seen as having completed the course), those who have had no prior experience in programming are more likely to have dropped out than those who had at least some prior experience on programming. Regarding age groups, those between 36 and 55 are most likely to continue at least to the 90% mark, while other groups – especially those between 18 and 25 – are more likely to abandon the course. Similarly, with respect to gender, we observe that the proportion of men increases, which means that women are more likely to drop out. Based on our analyses, this cannot be solely explained by prior experience in programming between these groups. Third, when considering who is likely to complete at least 10% of a subsequent course (i.e. to continue in another course in the series of courses), the trends observed in who are most likely to complete the first course become more dominant. In particular, prior experience in programming is a strong predictor of continuing in the course series, as is attending or having attended tertiary education.

These observations are problematic from multiple viewpoints. First, considering the government-level incentives where a part of the funding of the university comes from credits completed by non-degree students, there is considerable room for improvement. Second, similar to other open online courses in programming, it seems that although the course seems to attract a somewhat diverse population, those who continue in the course series are more likely to come from a similar mold. Considering the objective of the course, which is to teach principles of computer science for all, the "all" in this case could be argued to predominantly be middle-aged men who have at least some prior programming experience and who have at least attended tertiary education – unfortunately, this is a theme that has been observed also elsewhere [32].

As a part of our future work, we are reconstructing parts of the course materials with the objective of retaining the initial more diverse population. We are also looking beyond the demographic factors to study what within-course behaviors – such as material usage, behavior in programming exercises, and time-on-task – might be used to explain some of the dropouts in the light of the collected demographic data, and whether there is something that could be worked on there.

REFERENCES

[1] Nikos Benos and Stefania Zotou. 2014. Education and economic growth: A meta-regression analysis. *World Development* 64 (2014), 669–689.
[2] Inma Borrella, Sergio Caballero-Caballero, and Eva Ponce-Cueto. 2021. Taking action to reduce dropout in MOOCs: Tested interventions. *Computers & Education* 179 (2021), 104412.
[3] Abi Brooker, Linda Corrin, Paula De Barba, Jason Lodge, and Gregor Kennedy. 2018. A tale of two MOOCs: How student motivation and participation predict

learning outcomes in different MOOCs. *Australasian Journal of Educational Technology* 34, 1 (2018).

[4] Gayle Christensen, Andrew Steinmetz, Brandon Alcorn, Amy Bennett, Deirdre Woods, and Ezekiel Emanuel. 2013. The MOOC phenomenon: Who takes massive open online courses and why? *Available at SSRN 2350964* (2013).

[5] R Wes Crues, Genevieve M Henricks, Michelle Perry, Suma Bhat, Carolyn J Anderson, Najmuddin Shaik, and Lawrence Angrave. 2018. How do gender, learning goals, and forum participation predict persistence in a computer science MOOC? *ACM Transactions on Computing Education (TOCE)* 18, 4 (2018), 1–14.

[6] Paula G de Barba, Gregor E Kennedy, and Mary D Ainley. 2016. The role of students' motivation and participation in predicting performance in a MOOC. *Journal of Computer Assisted Learning* 32, 3 (2016), 218–231.

[7] Edward F Denison. 1966. Measuring the contribution of education to economic growth. In *The Economics of Education*. Springer, 202–260.

[8] Tawanna Dillahunt, Zengguang Wang, and Stephanie D Teasley. 2014. Democratizing higher education: Exploring MOOC use among those who cannot afford a formal education. *International Review of Research in Open and Distributed Learning* 15, 5 (2014), 177–196.

[9] Rodrigo Duran, Lassi Haaranen, and Arto Hellas. 2020. Gender differences in introductory programming: comparing MOOCs and local courses. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. 692–698.

[10] John Field. 2000. *Lifelong learning and the new educational order.* ERIC.

[11] John Field. 2011. Lifelong learning. In *Adult learning and education*, Kjell Rubenson (Ed.). Academic Press, 20–28.

[12] Nicoline Frølich, Evanthia Kalpazidou Schmidt, and Maria J Rosa. 2010. Funding systems for higher education and their impacts on institutional strategies and academia: A comparative perspective. *International Journal of Educational Management* (2010).

[13] Jeffrey A Greene, Christopher A Oswald, and Jeffrey Pomerantz. 2015. Predictors of retention and achievement in a massive open online course. *American Educational Research Journal* 52, 5 (2015), 925–955.

[14] Philip J Guo and Katharina Reinecke. 2014. Demographic differences in how students navigate through MOOCs. In *Proceedings of the first ACM conference on Learning@ scale conference*. 21–30.

[15] John D Hansen and Justin Reich. 2015. Democratizing education? Examining access and usage patterns in massive open online courses. *Science* 350, 6265 (2015), 1245–1248.

[16] Fredrik Heintz, Linda Mannila, Lars-Åke Nordén, Peter Parnes, and Björn Regnell. 2017. Introducing programming and digital competence in Swedish K-9 education. In *International Conference on Informatics in Schools: Situation, Evolution, and Perspectives*. Springer, 117–128.

[17] Arto Hellas, Petri Ihantola, Andrew Petersen, Vangel V Ajanovski, Mirela Gutica, Timo Hynninen, Antti Knutas, Juho Leinonen, Chris Messom, and Soohyun Nam Liao. 2018. Predicting academic performance: a systematic literature review. In *Proceedings companion of the 23rd annual ACM conference on innovation and technology in computer science education*. 175–199.

[18] Scott L Howell, R Dwight Laws, and Nathan K Lindsay. 2004. Reevaluating course completion in distance education: Avoiding the comparison between apples and oranges. *Quarterly Review of Distance Education* 5, 4 (2004), 243.

[19] Ben Jongbloed and Hans Vossensteyn. 2016. University funding and student funding: International comparisons. *Oxford Review of Economic Policy* 32, 4 (2016), 576–595.

[20] Ken Kay and Valerie Greenhill. 2011. Twenty-first century students need 21st century skills. In *Bringing schools into the 21st century*. Springer, 41–65.

[21] Bruce M King, Patrick J Rosopa, and Edward W Minium. 2018. *Statistical reasoning in the behavioral sciences.* John Wiley & Sons.

[22] René F Kizilcec, Chris Piech, and Emily Schneider. 2013. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *Proceedings of the third international conference on learning analytics and knowledge*. 170–179.

[23] Sophia Krause-Levy, Mia Minnes, Christine Alvarado, and Leo Porter. 2021. Experience Report: Designing Massive Open Online Computer Science Courses for Inclusion. In *Proceedings of the 26th ACM Conference on Innovation and Technology in Computer Science Education V. 1*. 95–101.

[24] Youngju Lee and Jaeho Choi. 2011. A review of online course dropout research: Implications for practice and future research. *Educational Technology Research and Development* 59, 5 (2011), 593–618.

[25] Juho Leinonen, Petri Ihantola, Antti Leinonen, Henrik Nygren, Jaakko Kurhila, Matti Luukkainen, and Arto Hellas. 2019. Admitting Students through an Open Online Course in Programming: A Multi-year Analysis of Study Success. In *Proceedings of the 2019 ACM Conference on International Computing Education Research*. 279–287.

[26] Yair Levy. 2007. Comparing dropouts and persistence in e-learning courses. *Computers & education* 48, 2 (2007), 185–204.

[27] Tharindu Rekha Liyanagunawardena, Andrew Alexandar Adams, and Shirley Ann Williams. 2013. MOOCs: A systematic study of the published literature 2008-2012. *International Review of Research in Open and Distributed Learning* 14, 3 (2013), 202–227.

[28] Piret Luik, Marina Lepp, Lidia Feklistova, Merilin Säde, Marili Rõõm, Tauno Palts, Reelika Suviste, and Eno Tõnisson. 2020. Programming MOOCs–different learners and different motivation. *International Journal of Lifelong Education* 39, 3 (2020), 305–318.

[29] Piret Luik, Reelika Suviste, Marina Lepp, Tauno Palts, Eno Tõnisson, Merilin Säde, and Kaspar Papli. 2019. What motivates enrolment in programming MOOCs? *British Journal of Educational Technology* 50, 1 (2019), 153–165.

[30] Andrew Luxton-Reilly. 2016. Learning to program is easy. In *Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education*. 284–289.

[31] Laura Pappano. 2012. The Year of the MOOC. *The New York Times* 2, 12 (2012), 2012.

[32] Justin Reich and José A Ruipérez-Valiente. 2019. The MOOC pivot. *Science* 363, 6423 (2019), 130–131.

[33] Charles A Schlosser and Mary Lagormarcino Anderson. 1994. *Distance education: review of the literature.* ERIC.

[34] Pablo Schoeffel, Raul Sidnei Wazlawick, and Vinicius Ramos. 2017. Impact of pre-university factors on the motivation and performance of undergraduate students in software engineering. In *2017 IEEE 30th Conference on Software Engineering Education and Training (CSEE&T)*. IEEE, 266–275.

[35] Duane F Shell, Leen-Kiat Soh, Abraham E Flanigan, and Markeya S Peteranetz. 2016. Students' initial course motivation and their achievement and retention in college CS1 courses. In *Proceedings of the 47th ACM technical symposium on computing science education*. 639–644.

[36] James H Steiger. 2004. Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological methods* 9, 2 (2004), 164.

[37] Steven F Tello. 2008. An analysis of student persistence in online education. In *Information communication technologies: Concepts, methodologies, tools, and applications*. IGI Global, 1163–1178.

[38] Tarmo Toikkanen and Teemu Leinonen. 2017. The code ABC MOOC: Experiences from a coding and computational thinking MOOC for Finnish primary school teachers. In *Emerging research, practice, and policy on computational thinking*. Springer, 239–248.

[39] European Union. 1995. 95/431/EC: Decision No 2493/95/EC of the European Parliament and of the Council of 23 October 1995 establishing 1996 as the 'European year of lifelong learning'. https://eur-lex.europa.eu/legal-content/PL/TXT/?uri=celex:31995D2493.

[40] Scientific United Nations Educational and Cultural Organization (UNESCO). 2016. Global education monitoring report summary 2016: education for people and planet: creating sustainable futures for all. (2016).

[41] Telle Whitney and Valerie Taylor. 2018. Increasing women and underrepresented minorities in computing: The landscape and what you can do. *Computer* 51, 10 (2018), 24–31.

[42] Robert Zemsky. 2014. With a MOOC MOOC here and a MOOC MOOC there, here a MOOC, there a MOOC, everywhere a MOOC MOOC. *The Journal of General Education* 63, 4 (2014), 237–243.