Self-efficacy Feedback Loops and Learning Experiences in CS1

Alex Lishinski
alishins@utk.edu
University of Tennessee
Knoxville, TN, USA

ABSTRACT

Self-efficacy refers to a students' beliefs about whether they can succeed in a particular domain. Students' self-efficacy beliefs are known to influence learning outcomes, through the effect they have on students' goal setting, learning strategies, and resilience behaviors. The strongest precursor to the formation of self-efficacy beliefs are students' own experiences completing learning tasks. Students complete learning exercises, they receive feedback, and they use this information to revise their self-beliefs. Successes can bolster an individual's self-efficacy for future learning tasks, and failures can damage an individual's self-efficacy. Self-efficacy can also form a reciprocal feedback loop, because performance feedback informs revisions to individuals' self-efficacy beliefs, and self-efficacy beliefs in turn influence adaptive behaviors that lead to better or worse learning outcomes. In this study we examined the self-efficacy beliefs of CS1 students at a large university during two semesters in an intensive longitudinal examination of the development of these beliefs. We examined CS1 students' self-efficacy beliefs and course performance over the course of a semester, using structural equation models designed to detect reciprocal effects. We found strong evidence in both semesters that such reciprocal feedback loops for self-efficacy can occur in CS1, although the reciprocal effects may die down by the end of the semester.

CCS CONCEPTS

• **Social and professional topics → CS1**.

KEYWORDS

CS1, Self-efficacy, Self-regulated learning

ACM Reference Format:
Alex Lishinski. 2023. Self-efficacy Feedback Loops and Learning Experiences in CS1. In *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1 (ITiCSE 2023), July 8–12, 2023, Turku, Finland*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3587102.3588837

1 INTRODUCTION

Self-efficacy refers to a students' beliefs about whether they can successfully perform the necessary actions to complete the task at hand [2]. Students' self-efficacy beliefs are known to strongly influence their learning outcomes, by virtue of the impact they have on students' learning strategies and resilience in the face of difficulties. Self-efficacy has been studied as a precursor to academic performance at every age level of student and across all subject areas and results consistently show a connection between self-efficacy and learning outcomes. The strongest precursor to the formation of self-efficacy beliefs is students' own experiences completing learning tasks. This unsurprising fact is predicted by the theory of self-efficacy as well as empirical work that has found this to be the case. When students complete a learning exercise, they receive feedback, both formally in what they get from instructors, and informally from the feedback that is implicit to the task, that lets them know how well they have done. It is easy to imagine a learning success bolstering an individual's self-efficacy for future learning tasks, and likewise a failure damaging an individual's self-efficacy. The theoretical model of self-efficacy suggests that they can form a reciprocal feedback loop, because performance feedback informs revisions to individuals' self-efficacy beliefs, and self-efficacy beliefs inform planning, goal-setting, and persistence in working through the task, all of which are related to student outcomes [2]. This aspect of the theoretical model of self-efficacy is a less examined aspect of self-efficacy, with a tiny number of studies involving self-efficacy examining this aspect of the construct [60, 63]. Some prior research in computer science education has implied the existence of a reciprocal feedback loop process in computing education [35], but none to date has examined this rigorously using a suitable model. To that end, this study addresses one primary research question.

- RQ1: Are there reciprocal effects between course performance and self-efficacy in CS1?

2 LITERATURE REVIEW

2.1 Self-efficacy

Self-efficacy [3], is one of the most studied education research motivational constructs. It refers to beliefs about one's ability to achieve success by performing the behaviors needed. Self-efficacy is important because it influences the amount of effort people will choose to expend to overcome difficulties. According to meta-analyses including tens of thousands of students, self-efficacy strongly influences student outcomes and persistence [8, 20, 47, 56].

Self-efficacy is part of the self-regulated learning (SRL) model of motivation. According this model, the learning process is an iterative cycle of forethought, performance, and self-reflection [40, 52]. Self-efficacy is connected to this cycle at every stage. It influences goals and planning in the forethought stage, influences attention focusing and learning strategies in the performance stage, and is revised in the self-reflection stage. there is also support for the notion that self-efficacy is not merely correlated with, but causally related to learning outcomes [13, 14]. This lends credibility to the notion of improving self-efficacy through pedagogical interventions.

ITiCSE '23, July 10-12, Turku, Finland

ACM ISBN 979-8-4007-0138-2/23/07…$15.00
https://doi.org/10.1145/3587102.3588837

Self-efficacy beliefs are connected with a process of development that is continuous and iterative. Students continuously judge their own performance based on whatever feedback or information they receive, and whatever inferences they make from that information. They then adjust their self-efficacy beliefs using those judgments. This can become a reciprocal feedback loop, because the revised self-efficacy beliefs will influence goal setting, learning strategies, persistence, and other self-regulated learning behaviors in future tasks in the domain. These behaviors will then themselves influence outcomes on the next task. In this way, self-efficacy can form a reciprocal feedback loop where poor performance leads to worse poor performance via the effect on self-efficacy. This corresponds to the theoretical framework of self-efficacy, and has been observed empirically as well [60, 63].

2.2 Self-efficacy in CS

Self-efficacy is context-specific. Therefore we must consider students' self-efficacy that is specific to computer science. Computing education researchers have been giving increasing attention to self-efficacy in recent years [36]. Self-efficacy has been found to relate to learning outcomes in computing courses, much like in other fields (e.g. [29, 43, 62]). Self-efficacy may be one of the most best predictors of success in CS in fact, with self-efficacy performing similarly to fine grained behavior based algorithms of students programming activities [61]. Interest in self-efficacy in computing education research has surged in the last several years, with researchers developing new CS self-efficacy instruments [5, 9, 54], incorporating self-efficacy into a student success prediction tool as a predictor [42], or using self-efficacy as an outcome to evaluate the impact of competitive enrollment policies [38].

Prior research has also indicated a need for greater self-efficacy support for women in CS. Women's self-efficacy is lower than that of men in STEM domains, CS in particular [4, 10, 11, 21]. Self-efficacy differences have also been connected to persistence above the influence of grades [22], so there is both a participation and a success outcome of self-efficacy to consider. These self-efficacy gender differences have also been show to persist beyond into the workplace as well, despite there being no gender differences in competence [32]. Self-efficacy has also been shown to be significantly related to career-related interests and choices across domains, so it is important to consider the impact of self-efficacy beyond immediate learning outcomes [30, 31, 37].

The development of self-efficacy in CS classes is complicated because self-efficacy develops in a continuous iterative process that is reciprocal with self-regulated learning behaviors, and which can create feedback loops, positive or negative, depending on the environment which may also differ by gender [35]. Students make self-assessments during their work that inform revisions to their self-efficacy beliefs. Qualitative research on self-efficacy beliefs of CS novices has identified a common sort of experience had by CS students that had a potentially large impact on self-efficacy judgments, the so-called "hit by lightning" experience [25–27]. This sort of experience occurs when students are compiling or running code. Students often begin with confidence and an expectation of success, only to be surprised with an unexpected error [25]. The hit by lightning experience often leaves students confused, frustrated, overwhelmed, annoyed, and with little sense of what to do next [25]. The strong emotional character of these experiences influences the ways that students modify their self-efficacy beliefs, particularly with novices [25]. If paired with an appropriate growth mindset, experiences of failure need not have negative self-efficacy consequences, but the way that computing work can produce these strong negative emotional responses in students makes it more likely that failures will induce negative revisions to their self-efficacy beliefs.

3 METHODS

3.1 Participants

This study reports on a study that took place over one academic year (two semesters) at a large university in the United States. The course being studied was a CS1 course, which served CS majors as well as non-majors and was taught in python. The data analyses presented in this paper comprise one component of a larger study. The fall semester iteration of the course included a total of over 600 students, of which 452 provided informed consent for the use of their data and usable repeated measures and covariate data. The spring semester included over 700 students, and 612 students provided informed consent and usable repeated measures and covariate data, including pretest and posttest measures, demographic and other background variables, ESM self-efficacy data, and CS1 course grade data.

The gender breakdown of the sample was 78% male and 22% female (pulled from registrar data) for both semesters. Race/ethnicity data was not collected for the first semester, but the second semester data showed that the population was 57% White, 11% Asian 5% Black 3% Hispanic 2% two or more races, and 22% Not Reported. The students in this study were from a variety of majors, with a substantial contingent of CS and engineering majors (61.3 % CS, Mech Eng, and Comp Eng combined), with a variety of other, mostly STEM majors (e.g. 4% Math, 2.8% Econ, 2.6% Stats, as well as many different business majors, each in small numbers).

3.2 Data Collection

As a part of the larger study that this data analysis was drawn from, a number of other survey measures and other covariate data were collected from students. Some of these were used in the multiple imputation models to generate imputed values for the repeated measures used in the primary models of interest. These covariates included four subscales from the Motivated Strategies for Learning Questionnaire (MSLQ). The MSLQ includes 83 items across 15 subscales, which cover constructs from two main areas: motivation and learning strategies [41]. Four MSLQ constructs were collected as a pretest at the beginning of the semester: self-efficacy, intrinsic goal orientation, extrinsic goal orientation, and metacognitive self-regulation. These scales were chosen due to the theoretical importance of these constructs in self-regulated learning, as well as the previously documented reliability of these particular subscales. The MSLQ self-efficacy subscale was also used in the substantive models presented in this paper. The validity and reliability of this scale have been well-established by previous research. Previous reliability values have been reported between .91 and .93, and correlations with same semester course grades for undergraduate students have been reported between .37 and .41 [8, 41]. The scale consists of eight Likert scale items. The items were assessed using a 7 point

Likert scale and sum scores were computed across the eight items. The full self-efficacy scale was collected at the beginning of the semester as a baseline covariate, whereas repeated measures data on self-efficacy was collected using a single item, described in greater detail below.

Scales corresponding to the Big Five personality traits (extraversion, agreeableness, conscientiousness, emotional stability, and openness to experience) were also collected from students, and were used as additional covariates for the multiple imputation models (described in greater detail in section 3.4 below). These were measured using the Big-Five factor markers developed by Goldberg [16], which included a total of 50 7-point Likert items for assessing individual trait values. These markers were made available through the international personality item pool (ipip.ori.org), a project developing and collecting sets of personality items that are made freely available for use by researchers [16]. These items have been analyzed for validity and reliability, and the available evidence suggests that they are able to meet these requirements [17, 18, 33, 45, 64]. The five personality trait scales each included ten 7-point Likert items. We also measured students' problem solving ability using a test based on the PISA problem solving exam. PISA (the Programme for International Student Assessment) is an international education study that tests 15 year old students on various subjects in dozens of countries.

The repeated measures data was collected 10 times per semester during both semester on four dimensions of students' experiences during the semester while working on programming projects. The dimension that is relevant to the results presented in this paper is self-efficacy. Self-efficacy was assessed in connection with each programming project (excluding the first, relatively trivial project) using the fourth item from the self-efficacy scale of the MSLQ: "Considering the difficulty of the course, the teacher, and my skills, I think I will do well in this class." The single item was chosen in order to assess students' self-efficacy on a repeated basis after each of the 10 programming projects, without having to require students to respond to all eight self-efficacy items each time.

This particular item was chosen from the self-efficacy scale because it had the highest correlation with the overall scale scores, or to use the terminology from factor analysis, it was the highest loading item. Using a single item, particularly the highest loading item, from a validated scale for repeated measures data collection is a common practice in intensive longitudinal studies [46, 51]. Prior research has shown that correlations with theoretically relevant variables are substantially similar whether one uses a full scale or a single item indicator and thus the answers to substantive questions are virtually identical, so this measurement approach has a strong case for having psychometric validity [7, 15]. The item was administered to all students 10 times over the semester by adding them to the instructions for programming projects 2-11, with students asked to insert their answer as a comment at the bottom of their project source code.

3.3 Reciprocal Effects Model

In order to examine whether there were reciprocal effects of self-efficacy on project outcomes, a special type of Structural Equation Model (SEM) was used: autoregressive cross-lagged models. This type of model was used to examine whether there were reciprocal relationships between self-efficacy and project score outcomes over the length of the course. The autoregressive cross-lagged model was suitable for this analysis because it allows one to examine reciprocal relationships between constructs over time [53]. An autoregressive cross-lagged model includes both autoregressive and cross-lagged relationships between variables. Autoregressive relationships are the effects of one variable on later iterations of that same variable, for example, the effects of self-efficacy on later self-efficacy. Cross-lagged effects are the effects of one construct on another construct across time points, for example, the effects of self-efficacy on later programming project scores. The model assumed that constructs affected one another across time points in order to capture the presumed reciprocal relationships between the constructs. By fitting this type of model, it could be determined whether there was evidence of reciprocal effects. Such evidence would support the notion that self-efficacy and project outcomes can operate as a feedback loop.

In order to fit the autoregressive cross-lagged model to the self-efficacy and project data, the data was condensed. This was necessary because of the recommended sample size guidelines for structural equation models, which suggest that 20 observations per estimated parameter is ideal [28]. Instead of considering the 10 time points separately, the variables were combined into three time chunks. The project scores and self-efficacy scores from projects 2,3, and 4, projects 5,6, and 7, and projects 8,9, and 10 were each averaged to create three project score measures and three measures of self-efficacy. The self-efficacy questions from project 11 were not included in the model, because of significantly greater non-response for that project due the lowest project score being dropped. The data was condensed in this way to reduce the number of parameters to be estimated in the model such that the available sample size was sufficient to reasonably fit the model.

The model consisted overall of 7 variables. Three aggregated time variables for programming project scores, and three aggregated self-efficacy repeated measures as described above. The seventh variable included was the full eight item self-efficacy scale given at the beginning of the semester, prior to any of the repeated measures data. This was included as a predictor of the initial project score and self-efficacy values, as a known predictor in the prior case, and as a baseline in the latter case.

The paths included in the model were those that make up the autoregressive cross-lagged model. Autoregressive paths were paths from earlier aggregated repeated measures to later ones, from variable 1 to variable 2, from variable 2 to variable 3, and from variable 1 to variable 3. These autoregressive paths were included for both project scores and repeated self-efficacy measures. The cross-lagged paths are paths from one side of the feedback loop to the other, moving forward one step in time. These paths were included from self-efficacy to programming projects, and from programming projects to self-efficacy, across all 3 aggregated variables. Finally, the model included a path from baseline self-efficacy to the first aggregated variable for both project scores and timepoints. The model specification is shown in figure 1.

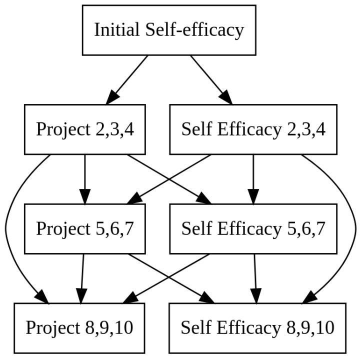

Figure 1: Auto-regressive Cross-lagged Model Diagram

3.4 Multiple Imputation

One methodological challenge that we encountered in this study was missing data. There was a substantial amount of missing data on repeated measures variables and simply fitting the models on the complete cases (listwise deletion) can lead to biased effect estimates, unless the missing data are missing completely at random (which is unlikely) [1]. A better approach in many cases is to try to account for the missing data with statistical techniques. One such technique is called multiple imputation (MI). The term imputation refers to simply filling in missing values of variables with a value so that data analysis can be conducted. This can be as simple as filling in the mean value of the non-missing cases for the missing cases, or it can involve using a model to generate estimated values based on available variables. Multiple imputation involves the latter approach, but repeated multiple times, generating multiple different imputed data sets. After the imputed data sets are generated, the model of interest is then fit to each data set and the estimated model parameters are pooled across these multiple iterations, with correction factors applied to correct standard errors and p-values for the effects of doing the multiple imputation process [57]. The biggest question for using MI is how many imputed data sets are required. Multiple imputation produces asymptotically unbiased parameter estimates as the number of imputations increases, but large numbers of imputations are often computationally intensive, especially for complex models [57].

The classical guidance on MI suggested 3-10 imputations as sufficient for most cases [50] but this recommendation has been updated over time as computational resources have allowed for more complex simulation studies of different imputation scenarios, and it has been realized that while lower numbers of imputations are sufficient for robust point estimates, standard error estimates often require more imputations [59]. The most recent guidance for the number of imputations is based on the fraction of missing information (FMI), which is a quantity that is estimated in an imputation run which indicates, not the amount of missing data, but rather the "proportion of variation in the parameter of interest due to the missing data" [19]. The formula for the sufficient number of imputations based on the FMI also depends on the tolerance that one has for inflation of the standard errors – more imputations means less inflation of the standard errors [58]; an often-used guideline for the acceptable level of inflation is <5% [49]. We estimated the FMI for our data at 0.447, which would indicate that we should use around 41 imputations for our data based on the formula presented by von Hippel [58]. We chose to use additional imputations to ensure the robustness of our estimates, so we used 200 imputations, which puts the inflation of our standard errors at approximately 2.8%. More imputations than this adds very little; a test run using 400 imputations produced virtually identical results to 200 imputations.

The data used for this study came from three main sources, the "pretest" set of surveys collected at the beginning of the semester, the repeated measures data collection in which students' self-efficacy was measured at 10 points throughout the semester, and the grade data from students' programming projects, of which there were 11 during the semester. The pretest data collection used validated scales and measures to collect data at the beginning of the semester on relevant student characteristics, including the data on the four MSLQ scales, the big five personality traits, and the problem solving pretest. These three sets of variables were all included in the imputation models.

The imputation was conducted on the repeated measures data prior to aggregating the variables into the three chunks of timepoint to ensure that each of these combined measures was created using the same number of pieces of information. We kept the two semesters of data separate because although the two semesters used the same basic structure, the projects differed accross the two semesters using different content as well as grading rubrics, so the data could not be combined due to these differences. The imputed datasets were generated using the mice package in R [57], which uses the chained equations algorithm to produce imputed data sets according to the fully conditional specification, where each variable with missing data is imputed with its own model that is appropriate to the type of variable. The imputations were generated using the random forest method, as a comparative analysis has indicated that this method generates superior imputations to other methods [23]. After creating 200 sets of imputations for the data from each semester, the aggregated variables were created, the autoregressive cross-lagged model was fit to each of the imputed sets of data using the lavaan R package [48], and the coefficient estimates were pooled and standard errors were calculated using Rubin's rules, using the semTools R package [24]. The analyses were conducted using R version 4.2.2 [55], and the path diagram in figure 1 was generated using the lavaanPlot package [34].

4 RESULTS

The autoregressive cross-lagged models were fit to the (imputed) data sets for both semesters. The model fit for both semesters was good. The test statistic and the RMSEA confidence interval indicate that the fit of the model to the fall semester data was quite good: (χ^2 (6) = 9.589, p > 0.05, RMSEA = 0.036, P-value RMSEA <= 0.05 = 0.655). (Note that good model fit is indicated by a χ^2 p-value greater than 0.05, and an RMSEA below 0.05) The test statistic and the RMSEA confidence interval indicate that the fit of the model to the spring semester data was also quite good, better in fact than that of the fall semester data: (χ^2 (6) = 5.313, p > 0.05, RMSEA = 0.000, P-value RMSEA <= 0.05 = 0.955).

The model coefficients for both semesters are shown in table 1. Most of the paths were statistically significant ($p<0.05$). The only exceptions were the path from baseline self-efficacy to the first set of project scores in the fall semester model, and the cross-lagged path from self-efficacy on projects 5,6, and 7 to the scores on projects 8,9,10 for the spring semester model. This suggests that the reciprocal relationship may weaken by the end of the semester. To summarize, consistent with the good overall fit of the model, all 12 autoregressive effects between the two models were significant paths, and 7 of the 8 cross-lagged effects were as well. Therefore, since the autoregressive cross-lagged models were a good fit to the data in both semesters, with the exception of one cross-lagged path, this suggests that self-efficacy and project scores significantly impacted themselves over time, not just from one time chunk to the next, but across the length of the course.

The estimated covariances provide the remainder of the picture, showing the concurrent relationship between the self-efficacy repeated measures and the project scores at the same time. The covariances between the three pairs of repeated measures variables is shown in table 2. The estimated covariances showed a pattern consistent with the path coefficients. Five of the six covariances were statistically significant, showing that the relationship holds over time. The sole exception to this was in the spring semester model. The final correlation between projects is not significant for the spring semester model, again suggesting that the reciprocal relationship between the two constructs weakens over time.

The autoregressive cross-lagged model was used to test for the evidence of reciprocal effects over time between students' self-efficacy and their programming performance. One of the eight cross-lagged effects failed to be statistically significant, suggesting that although there was overall strong evidence for reciprocal effects, they may taper off towards the end of the semester.

5 DISCUSSION

This study investigated the self-efficacy beliefs of CS1 students while working through programming projects, using repeated measures from across a CS1 course, in order to look for evidence of reciprocal effects with programming project scores. The more evidence of reciprocal effects that could be found (i.e., significant cross-lagged paths), the more evidence that there would be of a feedback loop process with students' self-efficacy and learning outcomes. The autoregressive cross-lagged path models tested whether there were such reciprocal effects by fitting a path model including cross-lagged paths between self-efficacy and project scores at different time points to model reciprocal effects, and autoregressive paths to control for prior values of self-efficacy and project scores.

The results showed substantial evidence of reciprocal effects, as seven of the eight cross-lagged paths were significant. This pattern of results suggested that there were reciprocal effects, but that these perhaps tapered off towards the end of the semester. One possible explanation for this may be that students are less prone to revise their self-efficacy beliefs in a reactionary way as they gain more competence with troublesome threshold concepts, which have been found to be a significant source of emotional difficulty for students [12]. This is consistent with previous research, which has found that emotionally difficult experiences in programming drive students' revisions of their self-efficacy beliefs [26].

The results of this study conform with prior work, which suggests that there should be reciprocal effects between experiences of success or failure and self-efficacy belief formation due to the nature of the construct [2, 39, 65, 66]. The limited empirical research on self-efficacy feedback loops has also suggested that self-efficacy may have such a relationship with success outcomes [6, 63], as well as related self-evaluation constructs like perceived goal achievement [60]. This particular study is limited to a single context and a single measurement approach, and these results may not generalize outside to these contexts. Given the evidence from previous research for self-efficacy reciprocal effects, along with the evidence for reciprocal effects of self-efficacy in this study, further research could investigate whether different measurement approaches to self-efficacy show different levels of reciprocal effects. Specifically, researchers could use programming-specific self-efficacy questions (e.g. [44]), assess self-efficacy in courses beyond CS1 [9], and attempt to determine which junctures of the course would be most significant for assessing self-efficacy.

The significant covariances and path coefficients in the path models support the hypothesis of a reciprocal relationship between self-efficacy. This suggests that if the programming projects can cause strong emotional reactions that influence self-efficacy, these reactions can create sustained effects over the course of the semester, although one of our data sets suggested that by the end of the semester the reciprocal effect had dissipated. For CS teachers this result implies that self-efficacy beliefs are something that should be supported because initial self-efficacy and negative experiences may have a lasting impact. Overall, the results of these analyses suggest that students' strong reactions to programming projects can create reciprocal feedback loops where self-efficacy beliefs impeded their future performance.

5.1 Limitations

This study is of course limited by the sample, which is not a random sample that would enable inferences to a national or international CS1 population. The population of study came from just one university and the students participating in this study were a self-selected subset of students in the CS1 course. This study is more of an exploratory investigation of students' CS1 experiences in one context, and future research is needed to determine whether the findings observed in this study would generalize to other contexts.

Other limitations of this study dovetail into directions for future research. We only looked at a small number of student level factors that could influence students experiences in CS1, but there are certainly many others at the classroom level, that would bear examination in future research. In the context of our analyses, these could be covariates in our imputation models, or additional components of a more complex iteration on our structural model. For example, we did not include in our analyses any measures of the context in these classes that might inform why students' self-efficacy developed the way we observed. These could be measures of the general classroom climate, the structure of the course activities, the features of the assignments themselves, and even teacher factors like instructor autonomy support, which have been shown to impact students'

Table 1: Model Coefficients Table

DV	IV	coef_FS	se_FS	t_FS	pval_FS	coef_SS	se_SS	t_SS	pval_SS
Projects 5,6,7	Projects 2,3,4	1.00	0.09	11.47	<0.001	0.92	0.08	11.08	<0.001
Projects 5,6,7	Self-efficacy 2,3,4	1.93	0.53	3.65	<0.001	0.98	0.35	2.78	0.005
Self-efficacy 5,6,7	Projects 2,3,4	0.05	0.01	3.92	<0.001	0.06	0.01	5.15	<0.001
Self-efficacy 5,6,7	Self-efficacy 2,3,4	0.94	0.08	11.27	<0.001	0.67	0.05	12.96	<0.001
Projects 8,9,10	Projects 5,6,7	0.46	0.06	7.19	<0.001	0.57	0.05	11.16	<0.001
Projects 8,9,10	Self-efficacy 5,6,7	0.79	0.34	2.29	0.022	0.14	0.30	0.48	0.632
Projects 8,9,10	Projects 2,3,4	0.39	0.11	3.67	<0.001	0.23	0.09	2.50	0.012
Self-efficacy 8,9,10	Projects 5,6,7	0.03	0.01	3.75	<0.001	0.02	0.01	3.46	<0.001
Self-efficacy 8,9,10	Self-efficacy 5,6,7	0.48	0.06	7.72	<0.001	0.53	0.04	12.58	<0.001
Self-efficacy 8,9,10	Self-efficacy 2,3,4	0.33	0.10	3.32	<0.001	0.14	0.05	2.85	0.004
Self-efficacy 2,3,4	Initial Self-efficacy	0.07	0.01	5.40	<0.001	0.13	0.02	8.48	<0.001
Projects 2,3,4	Initial Self-efficacy	0.13	0.08	1.58	0.115	0.25	0.07	3.47	<0.001

Table 2: Model Covariance Table

DV	IV	var_FS	se_FS	t_FS	pval_FS	var_SS	se_SS	t_SS	pval_SS
Projects 2,3,4	Self-efficacy 2,3,4	6.89	1.75	3.93	<0.001	7.64	1.77	4.31	<0.001
Self-efficacy 5,6,7	Projects 5,6,7	9.79	3.11	3.15	0.002	11.93	2.90	4.11	<0.001
Self-efficacy 8,9,10	Projects 8,9,10	7.53	3.03	2.49	0.013	2.66	2.24	1.19	0.235

motivational outcomes in prior research. There are certainly other student factors that would be worth examining in future research as well.

REFERENCES

[1] Paul D. Allison. 2001. *Missing Data*. Sage Publications.

[2] Albert Bandura. 1977. Self-efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological review* 84, 2 (mar 1977), 191–215. http://www.ncbi.nlm.nih.gov/pubmed/847061

[3] Albert Bandura. 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Prentice-Hall, Englewood Cliffs, NJ.

[4] Sylvia Beyer. 2014. Why are women underrepresented in Computer Science? Gender differences in stereotypes, self-efficacy, values, and interests and predictors of future CS course-taking and grades. *Computer Science Education* 24, 2-3 (2014), 153–192. https://doi.org/10.1080/08993408.2014.963363

[5] Jyoti Bhardwaj. 2017. In search of self-efficacy: development of a new instrument for first year Computer Science students. *Computer Science Education* (2017), 1–21. https://doi.org/10.1080/08993408.2017.1355522

[6] Gian Vittorio Caprara, Roberta Fida, Michele Vecchione, Giannetta Del Bove, Giovanni Maria Vecchio, Claudio Barbaranelli, and Albert Bandura. 2008. Longitudinal analysis of the role of perceived self-efficacy for self-regulated learning in academic continuance and achievement. *Journal of Educational Psychology* 100, 3 (2008), 525–534. https://doi.org/10.1037/0022-0663.100.3.525

[7] Felix Cheung and Richard E. Lucas. 2014. Assessing the validity of single-item life satisfaction measures: results from three large samples. *Quality of Life Research* 23, 10 (2014), 2809–2818. https://doi.org/10.1007/s11136-014-0726-4

[8] Marcus Credé and L Alison Phillips. 2011. A meta-analytic review of the Motivated Strategies for Learning Questionnaire. *Learning and Individual Differences* 21 (2011), 337–346. https://doi.org/10.1016/j.lindif.2011.03.002

[9] Holger Danielsiek, Laura Toma, and Jan Vahrenhold. 2017. An Instrument to Assess Self-Efficacy in Introductory Algorithms Courses. *Proceedings of the 2017 ACM Conference on International Computing Education Research* (2017), 217–225.

[10] Åge Diseth, Eivind Meland, and Hans Johan. 2014. Self-beliefs among students =: Grade level and gender differences in self-esteem, self-efficacy and implicit theories of intelligence. *Learning and Individual Differences* 35 (2014), 1–8.

[11] Rodrigo Duran, Lassi Haaranen, and Arto Hellas. 2020. Gender differences in introductory programming: Comparing moocs and local courses. *Annual Conference on Innovation and Technology in Computer Science Education, ITiCSE* (2020), 692–698. https://doi.org/10.1145/3328778.3366852

[12] Anna Eckerdal, Robert Mccartney, Jan Erik Moström, Kate Sanders, Lynda Thomas, and Carol Zander. 2007. From Limen to Lumen: Computing students in liminal spaces. *Proceedings of the third international workshop on Computing education research* (2007), 123–132.

[13] Viktor Gecas. 1989. The Social Psychology of Self-Efficacy. *Annual Review of Sociology* 15 (1989), 291–316. https://doi.org/10.1111/j.2044-8309.1982.tb00523.x

[14] Marilyn E. Gist. 1992. Self-Efficacy: A Theoretical Analysis of its Determinants and Malleability. *Academy of Management Review* 17, 2 (1992).

[15] Katarzyna Gogol, Martin Brunner, Thomas Goetz, Romain Martin, Sonja Ugen, Ulrich Keller, Antoine Fischbach, and Franzis Preckel. 2014. "My Questionnaire is Too Long!" The assessments of motivational-affective constructs with three-item and single-item measures. *Contemporary Educational Psychology* 39, 3 (2014), 188–205. https://doi.org/10.1016/j.cedpsych.2014.04.002

[16] Lewis R. Goldberg. 1992. The Development of Markers for the Big-Five Factor Structure. *Psychological Assessment* 4, 1 (1992), 26–42.

[17] Lewis R. Goldberg. 1999. A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several Five-Factor models. In *Personality Psychology in Europe (Vol 7)*, I. Merveilde, I.J. Deary, F. De Fruyt, and F. Ostendorf (Eds.). Tilburg University Press, 7–28.

[18] Alan J Gow, Martha C Whiteman, Alison Pattie, and Ian J Deary. 2005. Goldberg's 'IPIP' Big-Five factor markers: Internal consistency and concurrent validation in Scotland. *Personality and Individual Differences* 39 (2005), 317–329. https://doi.org/10.1016/j.paid.2005.01.011

[19] Martin W. Heymann and Iris Eekhout. 2019. *Applied Missing Data Analysis With SPSS and (R)Studio*. https://bookdown.org/mwheymans/bookmi/

[20] Toni Honicke and Jaclyn Broadbent. 2016. The influence of academic self-efficacy on academic performance: A systematic review. *Educational Research Review* 17 (2016), 63–84. https://doi.org/10.1016/j.edurev.2015.11.002

[21] Chiungjung Huang. 2013. Gender differences in academic self-efficacy: A meta-analysis. *European Journal of Psychology of Education* 28, 1 (2013), 1–35. https://doi.org/10.1007/s10212-011-0097-y arXiv:0507464v2 [arXiv:astro-ph]

[22] Mica Hutchison, Deborah K. Follman, Melissa Sumpter, and George M. Bodner. 2006. Factors Influencing the Self-Efficacy Beliefs of First-Year Engineering Students. *Journal of Engineering Education* 95, 1 (2006), 39. https://doi.org/10.1002/j.2168-9830.2006.tb00876.x

[23] Sebastian Jäger, Arndt Allhorn, and Felix Bießmann. 2021. A Benchmark for Data Imputation Methods. *Frontiers in Big Data* 4, July (2021), 1–16. https://doi.org/10.3389/fdata.2021.693674

[24] Terrence D. Jorgensen, Sunthud Pornprasertmanit, Alexander M. Schoemann, and Yves Roseel. 2022. semTools: Useful tools for structural equation modeling. https://cran.r-project.org/package=semTools

[25] Päivi Kinnunen and Beth Simon. 2010. Experiencing Programming Assignments in CS1: The Emotional Toll. *Proceedings of the Sixth international workshop on Computing education research (ICER '10)* (2010), 77–85. https://doi.org/10.1145/1839594.1839609

[26] Päivi Kinnunen and Beth Simon. 2011. CS Majors' Self-Efficacy Perceptions in CS1: Results in Light of Social Cognitive Theory. *Proceedings of the seventh international workshop on Computing education research - ICER '11* (2011), 19–26. https://doi.org/10.1145/2016911.2016917

[27] Päivi Kinnunen and Beth Simon. 2012. My program is ok – am I? Computing freshmen's experiences of doing programming assignments. *Computer Science Education* 22, 1 (mar 2012), 1–28. https://doi.org/10.1080/08993408.2012.655091

[28] Rex B. Kline. 2015. *Principles and practice of structural equation modeling.* (4th ed.). Guilford publications.

[29] Feng Yang Kuo, Wen Hsiung Wu, and Cathy Lin. 2013. An investigation of self-regulatory mechanisms in learning to program visual basic. *Journal of Educational Computing Research* 49, 2 (2013), 225–247. https://doi.org/10.2190/EC.49.2.f

[30] Robert W. Lent and Gail Hackett. 1987. Career self-efficacy: Empirical status and future directions. *Journal of Vocational Behavior* 30, 3 (1987), 347–382. https://doi.org/10.1016/0001-8791(87)90010-8

[31] Robert W. Lent and Gail Hackett. 1994. Sociocognitive mechanisms of personal agency in career development: Pantheoretical prospects. In *Convergence in career development theories: Implications for science and practice*, M.L. Savikas and Robert W. Lent (Eds.). CPP Books, Palo Alto, CA, 77–101.

[32] Matthew J. Liberatore and William P. Wagner. 2020. Gender, Performance, and Self-Efficacy: A Quasi-Experimental Field Study. *Journal of Computer Information Systems* 00, 00 (2020), 1–9. https://doi.org/10.1080/08874417.2020.1717397

[33] B C Lim and R E Ployhart. 2006. Assessing the convergent and discriminant validity of Goldberg's international personality item pool - A multitrait-multimethod examination. *Organizational Research Methods* 9, 1 (2006), 29–54. https://doi.org/10.1177/1094428105283193

[34] Alex Lishinski. 2022. lavaanPlot: Path diagrams for Lavaan models via DiagrammeR. https://lavaanplot.alexlishinski.com/

[35] Alex Lishinski, Aman Yadav, Jon Good, and Richard Enbody. 2016. Learning to Program: Gender Differences and Interactive Effects of Students' Motivation, Goals and Self-Efficacy on Performance. *Proceedings of the 12th Annual International ACM Conference on International Computing Education Research (ICER '16)* (2016).

[36] Lauri Malmi, Judy Sheard, Päivi Kinnunen, Simon, and Jane Sinclair. 2020. Theories and Models of Emotions, Attitudes, and Self-Efficacy in the Context of Programming Education. *ICER 2020 - Proceedings of the 2020 ACM Conference on International Computing Education Research* (2020), 36–47. https://doi.org/10.1145/3372782.3406279

[37] Karen D. Multon, Steven D. Brown, and Robert W. Lent. 1991. Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology* 38, 1 (1991), 30–38. https://doi.org/10.1037/0022-0167.38.1.30

[38] An Nguyen and Colleen M. Lewis. 2020. Competitive enrollment policies in computing departments negatively predict first-year students' sense of belonging, self-efficacy, and perception of department. *Annual Conference on Innovation and Technology in Computer Science Education, ITiCSE* (2020), 685–691. https://doi.org/10.1145/3328778.3366805

[39] F Pajares. 1996. Self-Efficacy Beliefs and Mathematical Problem-Solving of Gifted Students. *Contemporary educational psychology* 21, 4 (1996), 325–44. https://doi.org/S0361476X96900259[pii]

[40] Ernesto Panadero, Anders Jonsson, and Juan Botella. 2017. Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review* 22 (2017), 74–98. https://doi.org/10.1016/j.edurev.2017.08.004

[41] Paul R. Pintrich, David A.F. Smith, Teresa Garcia, and Wilbert J. McKeachie. 1993. Reliability and Predictive Validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement* 53, 3 (sep 1993), 801–813. https://doi.org/10.1177/0013164493053003024

[42] Keith Quille and Susan Bergin. 2019. CS1: how will they do? How can we help? A decade of research and practice. *Computer Science Education* 29, 2-3 (2019), 254–282. https://doi.org/10.1080/08993408.2019.1612679

[43] Vennila Ramalingam, Deborah LaBelle, and Susan Wiedenbeck. 2004. Self-efficacy and mental models in learning to program. *Proceedings of the 9th annual SIGCSE conference on innovation and technology in computer science education - ITiCSE '04* (2004), 171–175. https://doi.org/10.1145/1007996.1008042

[44] Vennila Ramalingam and Susan Wiedenbeck. 1999. Development and Validation of Scores on a Computer Programming Self-Efficacy Scale and Group Analyses of Novice Programmer Self-Efficacy. *Journal of Educational Computing Research* 19, 4 (1999), 367–381. https://doi.org/10.2190/C670-Y3C8-LTJ1-CT3P

[45] Beatrice Rammstedt, Lewis R Goldberg, and Ingwer Borg. 2010. The measurement equivalence of Big-Five factor markers for persons with different levels of education. *Journal of Research in Personality* 44, 1 (2010), 53–61. https://doi.org/10.1016/j.jrp.2009.10.005

[46] Lisa Respondek, Tina Seufert, and Ulrike E. Nett. 2019. Adding previous experiences to the person-situation debate of achievement emotions. *Contemporary Educational Psychology* 58, February (2019), 19–32. https://doi.org/10.1016/j.cedpsych.2019.02.004

[47] Steven B. Robbins, Kristy Lauver, Huy Le, Daniel Davis, Ronelle Langley, and Aaron Carlstrom. 2004. Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological bulletin* 130, 2 (2004), 261–288. https://doi.org/10.1037/0033-2909.130.2.261

[48] Yves Rosseel. 2012. {lavaan}: An {R} Package for Structural Equation Modeling. *Journal of Statistical Software* 48, 2 (2012), 1–36. http://www.jstatsoft.org/v48/i02/

[49] Patrick Royston. 2004. Multiple imputation of missing values: Update of ice. *Stata Journal* 4, 3 (2004), 227–241. https://doi.org/10.1177/1536867x0500500404

[50] Donald Rubin. 1987. *Multiple Imputation for Nonresponse in Surveys.* Wiley, New York.

[51] Jason D. Runyan, Brian N. Fry, Timothy A. Steenbergh, Nathan L. Arbuckle, Kristen Dunbar, and Erin E. Devers. 2019. Using experience sampling to examine links between compassion, eudaimonia, and pro-social behavior. *Journal of Personality* 87, 3 (2019), 690–701. https://doi.org/10.1111/jopy.12426

[52] Dale H Schunk and Peggy A Ertmer. 2000. Self-regulation and academic learning: Self-efficacy enhancing interventions. In *Handbook of self-regulation.* Elsevier, 631–649.

[53] James P Selig and Todd D Little. 2012. Autoregressive and Cross Lagged Panel Analysis for Longitudinal Data. *Handbook of Developmental Research Methods.* Chapter 12 (2012), 265–278.

[54] Phil Steinhorst, Andrew Petersen, and Jan Vahrenhold. 2020. Revisiting Self-Efficacy in Introductory Programming. *ICER 2020 - Proceedings of the 2020 ACM Conference on International Computing Education Research* (2020), 158–169. https://doi.org/10.1145/3372782.3406281

[55] R Core Team. 2022. R: A Language and Environment for Statistical Computing. https://www.r-project.org/

[56] Jeffrey C. Valentine, David L. Dubois, and Harris Cooper. 2004. The Relation Between Self-Beliefs and Academic Achievement: A Meta-Analytic Review. *Educational Psychologist* 39, July 2015 (2004), 111–133. https://doi.org/10.1207/s15326985ep3902

[57] Stef Van Buuren. 2018. *Flexible Imputation of Missing Data.* CRC Press.

[58] Paul T. von Hippel. 2018. How Many Imputations Do You Need? A Two-stage Calculation Using a Quadratic Rule. *Sociological Methods and Research* 49, 3 (2018), 699–718. https://doi.org/10.1177/0049124117747303

[59] Paul T. von Hippel. 2019. How many imputations do you need? https://statisticalhorizons.com/how-many-imputations

[60] Kristin Waschle, Anne Allgaier, Andreas Lachner, Siegfried Fink, and Matthias Nuckles. 2014. Procrastination and self-efficacy: Tracing vicious and virtuous circles in self-regulated learning. *Learning and Instruction* 29 (2014), 103–114. https://doi.org/10.1016/j.learninstruc.2013.09.005

[61] Christopher Watson, Frederick W. B. Li, and Jamie L. Godwin. 2014. No Tests Required: Comparing Traditional and Dynamic Predictors of Programming Success. *Proceedings of the 45th ACM Technical Symposium on Computer Science Education (SIGCSE '14)* (2014), 469–474.

[62] Susan Wiedenbeck. 2005. Factors affecting the success of non-majors in learning to program. *Proceedings of the 2005 international workshop on Computing education research - ICER '05* (2005), 13–24. https://doi.org/10.1145/1089786.1089788

[63] Trevor Williams and Kitty Williams. 2010. Self-efficacy and performance in mathematics: Reciprocal determinism in 33 nations. *Journal of Educational Psychology* 102, 2 (2010), 453–466. https://doi.org/10.1037/a0017271

[64] Lijun Zheng, Lewis R Goldberg, Yong Zheng, Yufang Zhao, Yonglong Tang, and Li Liu. 2008. Reliability and concurrent validation of the IPIP Big-Five factor markers in China: Consistencies in factor structure between Internet-obtained heterosexual and homosexual samples. *Personality and Individual Differences* 45 (2008), 649–654. https://doi.org/10.1016/j.paid.2008.07.009

[65] Barry J Zimmerman. 1990. Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Education Psychology* 82, 1 (1990), 51–59. https://doi.org/10.1037/0022-0663.82.1.51

[66] Barry J. Zimmerman and Rafael Risemberg. 1997. Becoming a Self-Regulated Writer: A Social Cognitive Perspective. *Contemporary Education al Psychology* 22 (1997), 73–101. https://doi.org/10.1006/ceps.1997.0919