All the Pieces Matter: The Relationship of Momentary Self-efficacy and Affective Experiences with CS1 Achievement and Interest in Computing

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

Joshua M. Rosenberg
University of Tennessee, Knoxville
Knoxville, Tennessee, USA
jmrosenberg@utk.edu

ABSTRACT

There are significant participation gaps in computing, and the way to address these participation gaps lies not simply in getting students from underrepresented groups into a CS1 classroom, but supporting students to pursue their interest in computing further beyond CS1. There are many factors that may influence students' pursuit of computing beyond introductory courses, including their sense that they can do what CS courses require of them (their self-efficacy) and positive emotional experiences in CS courses. When interest has been addressed in computing education, research has treated it mostly as an outcome of particular pedagogical approaches or curricula; what has not been studied is how students' longer-term interest develops through more granular experiences that students have as they begin to engage with computing. In this paper, we present the results of a study designed to investigate how students' interest in computing develops as a product of their momentary self-efficacy and affective experiences. Using a methodology that is relatively uncommon to computer science education—the experience sampling method, which involves frequently asking students brief, unobtrusive questions about their experiences—we surveyed CS1 students every week over the course of a semester to capture the nuances of their experiences. 74 CS1 students responded 14-18 times over the course of a semester about their self-efficacy, frustration, and situational interest. With this data, we used a multivariate, multi-level statistical model that allowed us to estimate how students' granular, momentary experiences (measured through the experience sampling method surveys) and initial interest, self-efficacy, and self-reported gender (measured through traditional surveys) relate to their longer-term interest and achievement in the course. We found that students' momentary experiences have a significant impact on their interest in computing and course outcomes, even controlling for the self-efficacy and interest students reported at the beginning of the semester. We also found significant gender differences in students' momentary experiences, however, these were reduced substantially when students' self-efficacy was added to the model, suggesting that gender gaps could instead be self-efficacy gaps. These results suggest that students' momentary experiences in CS1, how they experience the course week to week, have an impact on their longer-term interest and learning outcomes. Furthermore, we found that male and female students reported different experiences, suggesting that improving the CS1 experiences that students have could help to close gender-related participation gaps. In all, this study shows that the granular experiences students have in CS1 matter for key outcomes of interest to computing education researchers and educators and that the experience sampling method, more common in fields adjacent to computer science education, provides one way for researchers to integrate the experiences students have into our accounts of why students become interested in computing.

KEYWORDS

CS1, Experience Sampling Method, Self-efficacy, Interest, Frustration

ACM Reference Format:
Alex Lishinski and Joshua M. Rosenberg. 2021. All the Pieces Matter: The Relationship of Momentary Self-efficacy and Affective Experiences with CS1 Achievement and Interest in Computing. In *Proceedings of the 17th ACM Conference on International Computing Education Research (ICER 2021), August 16–19, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3446871.3469740

1 INTRODUCTION

It is important to the field of computing education research to understand why students become and stay interested in computing. It is important broadly as we seek to understand how to better engage students in computing, and it is particularly important as we continue to work on broadening participation in computing. Broadening participation has many aspects, but one particular focus of the computing education community in recent years has been on increasing the participation of female students in computing [33]. The first step of this process involves introducing female students to computing and sparking their initial interest. The logical next step is to provide a supportive environment that will encourage female students to develop a more persistent interest in computing and pursue their goals to the fullest extent possible [104]. Computing education research can do more to understand how we can provide that supportive environment.

Currently, we do not know enough about how students develop a robust interest in computing, and what we can do to support that development. In particular, we need a more robust understanding of the granular and incremental experiences students have day to day

ICER 2021, August 16–19, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-8326-4/21/08…$15.00
https://doi.org/10.1145/3446871.3469740

and week to week in CS classes that, in the aggregate, contribute to their persistent attitudes towards the field, positive or negative. We know that motivational beliefs and attitudes towards CS that students bring with them to courses can impact their outcomes [63, 94], but we know much less about the impact of students' experiences within courses. This is of particular importance in CS1 classes, because these students are often CS novices, and so this can be a critical early period when a small number of experiences can have an out-sized impact on students' eventual attitudes toward computing. It is important that we study students' experiences in CS on a more granular level so we can understand the underlying process of how students develop their interest in the field.

There are many factors that may be involved in developing an interest in CS, but examining the motivational and affective aspects of students' experiences is likely to provide important insights. There are certainly other lenses through which to view this problem, but educational psychology provides a relatively robust theoretical foundation for understanding interest [85]. This research identifies motivational and affective experiences as critically important in understanding how students' interest is developed in different subjects. In particular, this study examines the development of interest as it relates to students' experiences of self-efficacy and one particular type of affective experience, frustration. Both of these constructs are established in education research as important factors in understanding educational outcomes, and self-efficacy in particular has been repeatedly shown to be important in computing education. How they are related to the outcome of interest in particular is a question that needs to be explored further.

In order to study these constructs and understand their relationship to interest in CS, it is necessary to observe these aspects of students' experiences as frequently as possible to get a sufficiently detailed overall picture. To that end, it is necessary to make use of so-called intensive longitudinal research methods, in which frequent repeated measures are used to capture variation and relationships between constructs over time. This study used a particular intensive longitudinal method, the Experience Sampling Method (ESM) [40] to measure students' self-efficacy and affective experiences frequently over the course of the semester in two different undergraduate CS1 courses. ESM is based on frequent, repeated measures of constructs that tend to vary over time, like interest. In addition to being well-suited to the study of constructs that vary over time, an additional affordance of this type of research design is targeting the repeated measures to particular relevant moments. In this study, we targeted our ESM data collection *signals* to times that were relevant to students' CS experiences.

This study also gave particular attention to the role of self-efficacy. In particular, exploring how self-efficacy at the beginning of the semester helped explain variability in students' affective experiences and interest. Finally, we sought to examine the role of gender in these models, especially the role of gender relative to other characteristics of students that may matter as much (or more) than their gender—such as their initial self-efficacy beliefs. Given prior research results pointing to self-efficacy gaps in STEM fields and CS in particular, we sought to examine what gender differences existed in the processes being examined.

1.1 Purpose of study

Prior research in computing education has not thoroughly examined the details of the formation of interest. We do have research results about broader curricular programs that have been shown to lead to better interest for groups of students. What we do not know, however, is how more momentary, granular experiences can make the difference in supporting students in developing a robust, well-developed interest in the field. We do not know how much students' affective experiences and self-efficacy can fluctuate over time, and how these aggregate to influence levels of interest. We also do not know whether there are gender differences in any of these patterns that might be informative about how to create a more supportive environment in CS classes to better support the engagement and persistence of women. This study seeks to address these gaps in the prior research, and examine just how much these momentary experiences can influence interest in the field, as well as learning outcomes in CS1.

The central goals of this study, then, were to understand students' affective experiences and self-efficacy measured over the course of the semester, and how these related to their interest in CS and achievement at the end of the course. To that end, this study was focused around the following two research questions:

- RQ1: How do students experience CS1 and what student characteristics and attitudes are related to their experiences?
- RQ2: How do their momentary experiences relate to changes in their interest and end-of-course achievement?

2 LITERATURE REVIEW

2.1 Interest as a construct

Interest is a deceptively simple construct. Everyone has an intuitive sense of what it means to be interested in something. However, when it comes to defining this construct theoretically and using it in research contexts, the research literature has produced a much more nuanced and complex picture of interest and the ways it develops. Educational psychology researchers have created several conceptual frameworks for interest, focusing on the different elements that comprise it [85]. These theories alternatively focus on interest as a set of *developmental stages* [41], as an *emotional state* [3], and as beliefs about the value of a potential subject or object of interest [89]. Because interest promotes self-regulated learning behaviors like persistence, conscientiousness, and engagement, supporting the development of interest makes it more likely that all learners can reach their potential [86].

All of these models for interest are potentially useful, but the four-phase model of interest development from Hidi and Renninger [41] has a particular appeal to help conceptualize how students go from initial exposure to a field like CS, to developing a sustained interest of the sort that leads to pursuing a degree and/or career in CS. According to Hidi and Renninger's [41] model there are four phases in the development of interest. The first is called *triggered situational interest*, which is a temporary psychological state of increased attention and engagement that can be sparked by characteristics of the task or situation, which is usually supported externally, and which may lead to a disposition to re-engage in a certain domain as interest develops further, but which also may not lead to further development of interest. The second phase is *maintained situational interest.* This phase is characterized by the psychological state of interest initiated in an episode of triggered situational interest being maintained over an extended period of time and/or recurring. Once again, the occurrence of an instance of maintained situational interest may lead to the further development of interest or it may not. The third phase of interest development is called *emerging individual interest*, which is a combination of the psychological state described by the first two phases, and the beginning stages of an enduring predisposition to engage in the domain. One characteristic aspect of emerging individual interest is that it is becoming largely self-generated and self-sustaining, although external support from peers and mentors can better support its development, because once again this stage of interest development may or may not lead to the next stage. The final stage is *well-developed individual interest*, which continues to manifest both as a psychological state and an enduring predisposition towards engagement with a particular domain. This stage is distinguished from emerging individual interest by greater positive feelings, stored knowledge and stored value about the domain of interest, and self-generation of ongoing engagement with the domain.

Hidi and Renninger's [41] four-phase model is particularly useful for understanding how students' interest in CS can develop at multiple levels and phases, from engaging students at the initial point of contact such that triggered situational interest in CS is generated, to supporting them to the point where they can reach the level of a well-developed individual interest in CS if they so desire. At the outset of learning CS in a CS1 class, it is conceivable that students could move from phase 1 to phase 3 if the environment is adequately supportive. How CS educators can best do this is another open question, but it is an important one, particularly as we seek to broaden participation in CS of underrepresented groups, because our goal should not be to broaden participation just at the level of phase 2 of interest, but all the way through phase 4.

The concept of interest has commonly been invoked in prior computing education research, but, while informative, this is most often done informally (e.g. [14, 27]), and examples of authors anchoring research to an established theoretical framework for interest is less common, but increasing in recent years. Interest in CS often appears in research as an outcome, such that the impact of a pedagogical approach or outreach effort can be evaluated. Prior research has investigated specific computing curricula that have been found to successfully increase students' interest. For example, work by Guzdial [38] has examined a media computation based approach to introductory computing, and work by Kafai and colleagues [47, 48] has examined the use of e-textiles into secondary level computer science classes. Some computing education studies have also invoked theoretical models of interest. For example, a recent study by McKlin et al. [69] used interest, based on Hidi and Renninger's four-phase model, to study the impact of combining music with CS on the interest of middle school students. Interest is also invoked in research investigating the underlying factors in broader trends in enrollment in CS, particularly related to the participation gaps. Relatively early work by Margolis et al. [67] examined the nuances of women's interest in computer science and some of the challenges they tend to face in pursuing and further developing their interest in the field. A well-known study by Cheryan et al. [17] showed how stereotypical cues contribute to lower levels of CS interest in female undergraduate students. Other studies have looked further at higher level influences on interest for women and other minoritized groups [16, 24, 49, 68], but what has not been examined is the influence of finer-grained experiences on the development of interest. For this reason, we focus on interest at a fine-grained level, as reported by students via ESM—primarily following prompts sent out after students completed key assignments for their course. In this way, the interest we study is more *situational* in nature. Thus, we refer to this fine-grained interest as situational interest—the kind referred to by Hidi and Renninger [41].

2.2 Self-efficacy

Self-efficacy is a construct first articulated by Bandura [6], and which is one of the most important (and studied) motivational constructs in education research. It refers to one's beliefs about one's abilities to perform the behaviors needed to achieve success. The importance of self-efficacy comes from the fact that it influences the amount of effort people are willing to expend to overcome difficulties, which explains why meta-analyses looking at tens of thousands of students in total have shown that self-efficacy is strongly related to student outcomes and persistence [19, 42, 72, 87, 99].

Self-efficacy is a component of the larger self-regulated learning (SRL) model that explains the learning process as an iterative cycle of forethought, performance, and self-reflection [77, 90]. Self-efficacy is connected to each stage of this cycle, influencing goals and planning in the forethought stage, influencing attention focusing and learning strategies in the performance stage, and being revised in the self-reflection stage. Interest connects to the SRL model as well, although the nature of this connection is subject to debate. Some research has suggested that self-efficacy and interest are two related but separate pathways to greater levels of self-regulated learning behaviors [57]. Prior research has further suggested that self-efficacy is not merely correlated with learning outcomes, but causally related [29, 30], suggesting that self-efficacy can be improved through pedagogical interventions. There are some general approaches to improving self-efficacy that are supported by research but investigations of implementing these in CS classes are uncommon. These approaches include growth mindset training [15], success attribution retraining [9, 97], peer modeling instruction [2, 91], goal setting instruction [91, 105], and self-assessment [5, 77].

Self-efficacy beliefs are the product of a continuous and iterative developmental process. Students, particularly novices, continually judge their own performance and make adjustments to their self-efficacy beliefs based on those judgments. This can result in a reciprocal feedback loop process, because the revised self-efficacy beliefs will influence goal setting, learning strategies, and persistence, among other self-regulated learning behaviors, which then in turn influence outcomes on the next task undertaken. It is in this way that self-efficacy can form a reciprocal feedback loop where poor performance leads to worse poor performance via the effect on self-efficacy. This fits the theory of self-efficacy and has been observed empirically as well [100, 103].

2.3 Self-efficacy in CS

As self-efficacy is a context-specific construct, it is sensible to consider students' particular computing self-efficacy. Self-efficacy has been noted as an important construct by computing education researchers for years, but researchers are paying increasing attention to it in recent years [66]. As in other fields, self-efficacy has been found to consistently relate to outcomes in computing courses (e.g. [56, 84, 102]). Research has further suggested that it may be one of the most important predictors of success in CS, with self-efficacy performing similarly to fine grained behavior based algorithms of students programming activities [101]. Particularly in the last few years there has been a surge of interest in self-efficacy in computing education research, with researchers developing new instruments for measuring self-efficacy in CS [8, 21, 96], as well as incorporating self-efficacy into studies in a variety of ways, such as in a student success prediction tool as a predictor [83], or as an outcome to evaluate the impact of competitive enrollment policies [75]. Some studies have even begun connecting pedagogical strategies to increases in students' self-efficacy. For example, a study by Peteranetz et al. [81] found that using computational creativity exercises led to higher self-efficacy in undergraduate computing students.

Furthermore, prior research has pointed to a need for greater support of self-efficacy for women in CS. Women have lower average self-efficacy in STEM domains than men, CS in particular [7, 23, 25, 43], and self-efficacy differences are linked to persistence above and beyond the influence of grades [44]. Research has also shown that these self-efficacy gender differences can persist beyond the educational environment and into the workplace, despite there being no gender differences in competence [60]. Self-efficacy has also long since 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 [58, 59, 72]. Recent efforts have made significant progress towards closing the CS gender gap, but continuing to build representation of women in the field, and more than that, putting women in a position to make an impact in the field, will need to involve building more self-efficacy supporting environments in computing education.

The development of self-efficacy in CS classes is more complicated than it appears for a few reasons. For one, 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 [65]. Another complicating aspect of self-efficacy in CS is that students make self-assessments that inform revisions to their self-efficacy beliefs. Qualitative research on students' formation of self-efficacy beliefs by Kinnunen and Simon [52–54] found that there was a common sort of experience had by CS students that had a potentially large impact on self-efficacy judgments. They called this the "hit by lightning" experience, which is the experience that students have when compiling or running code with confidence and an expectation of success, only to be surprised with an unexpected error [52]. The hit by lightning experience often left students confused, frustrated, overwhelmed, annoyed, and dazed with little sense of what to do next [52]. The strong emotional character of these experiences has a significant impact on the ways that students adjust their self-efficacy beliefs, which Kinnunen and Simon [52] also noted as a common through line of novices' programming experiences. 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 negatively impact self-efficacy.

Another challenging aspect of supporting students' self-efficacy in CS is that the ways that CS students self-assess are frequently problematic. Recent research has suggested that CS students' judgments of their own ability are both different in CS than in other fields, as well as inaccurate and not conducive to optimal learning. A recent study of CS1 students found that they often make self-judgments with a self-critical bias, judging that difficulties that occur while programming reflect more negatively on their own ability than for a hypothetical other student [35]. The way that students make these judgments are not only skewed negative, but they often are based on experiences that are perceived by students to reflect negatively on their ability, but which experts in programming experience too, such as planning, getting errors, and asking for help [34]. So students do not know how to make accurate self-assessments to inform their self-efficacy beliefs, and these judgments often have an overly negative valence. Combined with the frequency that CS students encounter emotionally difficult situations in doing programming, these findings make a compelling case that supporting students' self-efficacy in CS courses is particularly important.

2.4 Affective experiences

Affective (or emotional) experiences matter in education because they influence students' behaviors in a variety of complex ways [61, 78]. The concept of cognitive load is often invoked in computing education research, but the parallel notion of *affective load* is at least as important for researchers to consider, as "affective behavior initiates, maintains, and terminates cognitive behavior" [74]. This is particularly important in a problem solving centric domain like computer science, because affective experiences play a significant role in the self-regulated learning behaviors of students engaged in problem solving, influencing the focus of attention and biasing cognitive processes [39].

Affective experiences are also connected to students' interest—both theoretically and empirically [78]. The work by Ainley and Ainley [4] on the construct of interest has conceptualized interest as an affective state that can support learning as a key component of engagement [85]. This differs from Hidi and Renninger's [41] four-phase model of interest, in which affective experiences can be one precursor to the formation of interest, but regardless of where the theoretical lines are drawn it is clear that interest is connected to affective experiences. Given that interest development promotes many aspects of the cycle of engagement and learning, it is important that we pay attention to the affective dimension of interest formation.

Prior work in computing education has examined affective experiences from a variety of perspectives [66]. The qualitative work by Kinnunen and Simon [52–54] discussed above looked at a number of aspects of novice students' experiences in programming courses, looking both at self-efficacy beliefs and how students revised them, as well as the affective character of their experiences that informed how students were likely to make these modifications. Eckerdal et al., [26] examined emotional responses in connection with threshold concepts, those which are both difficult to understand and central to the discipline, and found that these concepts produced strong negative emotional reactions such as frustration and depression which gave way to excitement and euphoria once the concepts were understood. They emphasized that it was important to recognize these experiences and provide a supportive environment so that students are able to persevere through these difficult experiences. Other research has corroborated the hypothesis that affective experiences significantly impact outcomes for programmers. A study by Graziotin et al. [37] found a significant relationship between emotional valence and self-assessed productivity over a single 90 minute programming episode, and this held true for students as well as professional programmers. A connection between the emotions first year undergraduate students experience in class and their course outcomes was found by Ruiz et al. [88]. Bosch and D'Mello [10, 11] examined sequences of affective states in novice programming students, finding that the students would frequently transition from a state of flow to confusion to frustration, and back and forth between these states during a single programming episode. The frequency of these transitions was related to outcomes, but was also related to the type of scaffolding provided to students, suggesting that the learning environment can mediate the impact of affective states on learning outcomes. They also found significant patterns of co-occuring affective states, such as confusion with frustration, and curiosity with engagement.

Of the many types of affective experience that students can and do encounter in computer science, students' experiences of frustration are particularly salient. Frustration is the affective experience of central interest in this study because previous research on affective experiences in problem solving has tended to key in on frustration as one of the most salient emotions experienced [22, 39]. Furthermore, frustration has been linked more broadly to self-regulated learning behaviors [45, 71]. The research in computing education on emotions has likewise highlighted a central role for frustration [11, 52]. Frustration is certainly not the only type of affective experience that impacts computer science students' outcomes, but whether looking through the lens of problem solving or self-regulated learning, it is clear that frustration is of central importance in CS students' experiences.

How students' affective experiences may differ by their gender is a surprisingly less studied topic in computing education research. One study investigating students' affective experiences in CS1 looked at differences by gender on feelings of frustration, inadequacy, and pride, finding significant gender differences on the two negatively valenced emotions, as well as significant relationships between these experiences and course outcomes [64]. Gender differences in affective experiences have been investigated more extensively in other domains outside computing. Like self-efficacy, affective experiences have the capacity to create feedback loops as they are both causes and effects of academic achievement experiences [79]. Research has found that female students are more likely than male students to have negative affective experiences in STEM courses, which is especially concerning given the occurrence of reciprocal feedback loops [31, 73, 80]. For this reason, students' affective experiences in CS classrooms are a dimension that needs to be further addressed to support women in computing.

Prior research in computing education has used a variety of creative approaches, from automatic detection of students' emotions based on facial cues [36], to retrospective review by the students themselves of a videotaped programming session [11]. Another viable way to investigate affective experiences involves using surveys. As we discuss further in our method section, the experience sampling method is an intensive longitudinal research design that involves frequent measurement via survey over a constrained period of time. This technique is particularly well-suited to investigate constructs that change frequently and in response to different momentary contexts, which is why it is a frequently used method to study affective experiences (e.g. [51, 92, 95]).

3 METHOD

3.1 Context and participants

This paper reports on findings from a study in the fall, 2020, involving 74 undergraduate students enrolled in one of two CS1 courses at a large, public university in the United States. The two classes, which we refer to as CS1A and CS1B are the two CS1 courses offered at this university. CS1A is the traditionally offered CS1 course, taught in C++ and serving as an introduction to the computing course sequence for an audience of primarily students in the college of engineering, including computer science majors as well as other engineering majors required to take the course. CS1B is a newly created course, launched Fall semester 2020 and taught in Python, which teaches a broader introduction to problem-solving using programming that serves more as a standalone course for students who would not typically take CS1A. Students' gender as used in the following analyses was self-reported, with 22 students (29.7%) identifying as female, 50 (67.6%) identifying as male, and 2 (2.7%) choosing not to respond (an open-ended response option was offered but not used).

3.2 The Experience Sampling Method

ESM is an intensive longitudinal data collection method that is used to collect information about the daily life of individuals as it is perceived in the moment [40]. Concretely, ESM studies involve surveying participants frequently over a period of time, targeting the surveys to specific events to focus data collection on a specific area of interest or surveying at random to get an unbiased sampling of activities during the period. The characteristics of ESM allow it to combine the ecological validity of naturalistic observation, the non-intrusive nature of diaries, and the precision of scaled survey items [40]. These advantages allow researchers to answer questions with ESM that would not be answerable with standard survey research, and to do so at a scale that is not typically feasible with qualitative studies.

As with all research methods, ESM has strengths as well as weaknesses. The strengths of increased ecological validity, the ability to study within-person processes, and avoiding biases of traditional self-reports, such as memory bias, must be considered against potential downsides of ESM research [93]. These include self-selection bias, and respondent fatigue causing data quality issues [93]. The former is a limitation of this study that we note in the discussion section of this paper. However, this limitation came into play mostly due to the traditional survey research components of the study, as virtually all of the students willing to complete the traditional survey at the beginning of the semester went on to participate in the ESM data collection, so it is not unique to ESM. The issue of respondent fatigue is one that we gave great consideration to, and we developed a novel data collection approach, described in the following section, largely in order to minimize the burden on students and minimize this issue. Furthermore, our ESM data collection surveyed students typically just one time per week, which seemed to provide minimal risk of respondent fatigue that would lead to data quality issues. Some ESM studies can survey participants as often as seven to ten times per day, and empirical examinations of such data have provided good evidence for their validity and reliability [20].

ESM is used in many different research areas, and its popularity in education research has been growing in recent years [70, 106]. It is particularly useful for investigating student characteristics that change frequently over time, for example, emotions and engagement are frequent topics of study in ESM education studies. In this study, we used ESM to survey students more than once per week about their self-efficacy in CS1, their affective experience of frustration, and their momentary interest in CS1.

3.3 Data Collection

Against the many benefits provided by ESM must be weighed the burden upon students, as ESM studies place a larger than normal burden upon students. In order to minimize this burden, we pursued a novel approach to ESM data collection for this study. There are many means by which to conduct an ESM data collection, and these methods have evolved over time with technology. In the early days of ESM, it was common for surveys to be done on paper, with the signal to complete the survey given by a watch or beeper provided by the researchers [40]. Later, with the further development of portable computing devices, participants in ESM studies would often be given special portable devices that would signal and allow participants to input their survey responses. Most recently, with the ubiquity of smartphones, it has been most common for ESM participants to use an ESM app which handles signalling and data collection through a survey interface. This is a convenient and effective approach that reduces many of the drawbacks of earlier ESM data collection methods, however, we saw drawbacks to this approach that could be reduced further.

One goal we had for our use of ESM was to use this method in a way that minimized the burden on students in responding. We also wanted to build trust with our students. For these reasons, asking students to download an unfamiliar smartphone app with an unfamiliar interface was not ideal. Furthermore, we thought that the use of an application that students would only likely use for the purposes of this study could subtly reduce the naturalistic qualities of our data collection. Accordingly, we developed our own approach to ESM data collection based on sending (and receiving) SMS messages, or text messages.

To do this, we created a web application, which we first used during the spring semester 2020 (the semester before the present study) that surveyed students via SMS at specific times—primarily immediately after they ended their CS1 class sessions. The app would prompt students that it was time to begin the survey, and then ask them to answer 4-5 Likert scale items on a 1-5 scale. By leveraging an existing interface on students' phones, namely the SMS application, that students were already familiar with using their everyday life, we believe that we were able to reduce the burden to students in responding to our items, as well as more seamlessly integrate our ESM surveys into their everyday life. The approach worked well for data collection during the spring, 2020 semester (even during the COVID-19 pandemic, our ESM response rates approached 80%, and average survey response times were just over 20 seconds [62].

This approach worked well for the first semester when we had a reliable class schedule based on in person class meetings, but the COVID-19 pandemic and the subsequent shift to online asynchronous instruction demanded a shift in our approach. Instead of sending messages based on a fixed time schedule, which no longer allowed us to target our data collection to moments when we knew that students were going to be engaged with CS1, we changed to a more flexible and dynamic event contingent data collection plan. Data collection was now pinned to significant assignments in the course. Making use of notifications from the course LMS, students were prompted to respond to the ESM surveys by our app within 5 minutes of submitting assignments. So even though students did not do any class activities synchronously, we were able to reach out to each student at an isomorphic set of instances across the semester. These assignment linked prompts were sent 13 times for CS1A and 9 times for CS1B due to different numbers of assignments between the two courses. Additionally, we sent 5 non assignment linked ESM surveys to all students on a fixed time schedule every 2 weeks during the semester to establish a stable baseline of individual response tendencies. All surveys were framed the same way, as pertaining to their feelings towards their CS1 course as a whole at the moment surveyed.

3.4 Measures

Data collection for this study consisted of a pre-survey at the beginning of the course, a post-survey at the end of the course, and the regular ESM surveys during the semester. The pre-survey contained measures for students' self-efficacy and task value, the latter of which includes a construct for students' interest. The self-efficacy scale used was the 8 item Self-Efficacy for Learning and Performance scale from the Motivated Strategies for Learning Questionnaire [82]. This scale asks students about their self-efficacy relative to a specific course, but does not specifically address computer science content in the items (e.g. I'm confident that I can understand the basic concepts taught in this course). The task-value related belief scales were those presented by Conley [18], replacing "math" in the items with "computer science." We refer to these as measures of *initial self-efficacy* and *intial interest*. The post-survey included the same task-value measure (including the measure of interest); we refer to this measure as students' *end-of-course interest*. We refer to both of these interest measures—and the change over time from initial interest to end-of-course interest—simply as *interest*, distinct from *situational interest* as measured via an ESM item, as we describe next.

There were three items used in the ESM surveys administered to students via SMS message, one each targeting momentary experiences of *frustration*, *self-efficacy*, and *situational interest*. The items were contextualized with the following message: "Please indicate your agreement at this moment with the following statements about your experiences in CS1 on a 1-5 scale, with 1 indicating strong disagreement, 3 indicating that you neither agree nor disagree, and 5 indicating strong agreement." The statements then followed: "I feel frustrated", "I feel confident about being able to do the work going forward" (self-efficacy), and "I feel interested in computer science" (situational interest). The item for frustration followed a straightforward structure commonly used in ESM studies of affective experiences (e.g. [50]). The item targeting self-efficacy was adapted from the MSLQ self-efficacy scale for the context of the ESM survey. We removed the "in this class" language that is found in the MSLQ self-efficacy items as that context is provided instead by the survey prompt. Furthermore we combined the specific content of the three MSLQ items explicitly addressing feeling "confident" which respectively address students' confidence about "basic concepts", "the most complex material", and "assignments and tests" which we covered with the catchall phrase "the work." The interest item also follows a structure common to other ESM studies on domain interest.

3.5 Data analysis: Multivariate, multilevel models

Our research questions targeted three different aspects of students' experiences: a) the experiences students have, b) what things relate to those experiences, and c) how students' experiences impact course-long outcomes.

To analyze our data to answer these questions, we used multivariate, multilevel models. These models are extensions of linear regression models in two ways. First, these models allow for multiple outcome variables to be specified at once. In this way, both students' experiences of frustration and their interest, for example, can be understood through the use of the same model. In addition, these models are multilevel, in that they allow for the grouping structures present in the data to be reflected in the analysis. There were two such dependencies in our data: the dependencies associated with each student's responses, which share some degree of common variability on the basis of being from the same student (e.g., students who tend to experience interest in characteristic ways), and dependencies associated with particular assignments (e.g., some assignments generally provoke more frustrating experiences than others). To estimate these multivariate, multilevel models, we used the brms R package [12]. Because of the complexity of estimating such models, brms uses Markov Chain Monte Carlo (MCMC) estimation, rather than maximum likelihood. We estimated five models, the first three to answer RQ1, and the last two to answer RQ2:

1. First, we estimated a model with only the grouping factors–a null, or variance components model, intended to reveal the sources of variability in students' momentary experiences. In this model these sources include student-level, assignment-level, and residual (unexplained) variability.
2. Then, we estimated a model for the three momentary student experiences: their interest, frustration, and self-efficacy, using the same grouping factors from the first model, and adding gender, as well as an indicator for the course (for whether students were in CS1A or CS1B).
3. Next, we added students' initial self-efficacy and initial interest in CS to the second model, while keeping all of the other factors.
4. Finally, we added the course-long outcome of students' end-of-course interest with the same predictors for that outcome as for the three momentary experience outcomes (situational interest, frustration, and self-efficacy).
5. We also estimated a model identical to the fourth model, except replacing end-of-course interest with students' final course grade (as a percentage).

To facilitate the interpretation of the estimates, we standardized all continuous variables to have the properties $M = 0$, $SD = 1$. This includes the Likert data from the ESM data collection. Although Likert data is technically ordinal, which would make treating it as continuous inappropriate, there is significant evidence that statistical models are quite robust to treating Likert data as an ordinal approximation of a continuous variable, particularly with five or more scale points [76, 98], although we must note that this is still a matter of some contention amongst researchers [13]. As a side note, treating Likert data in this manner is also typical practice for ESM studies, (e.g. [28, 32, 46]). These transformations mean that model coefficients can be interpreted in SD units. A coefficient value means that a one-*SD* change in the independent variable equates to that coefficient value of a change in the dependent variable in *SD* units.

Interpretation of these models is comparable to but in a few ways different from models estimated using maximum likelihood methods. When using this MCMC estimation method, it is common to report the range around the estimate via a technique analogous to the use of confidence intervals, the use of so-called credible intervals [55]; we used 95% credible intervals. Finally, because the parameters are estimated as a distribution, the traditional null hypothesis statistical testing requires a change; instead of asking how likely an estimate is under the conditions of the null hypothesis (most often that the estimate is equal to zero), models using MCMC often involve calculating how much of the distribution for the estimate overlaps not with zero, but with a region of practical equivalence, or a ROPE [55]. We followed the guidelines from [55] for how to set the parameters for the MCMC estimation and how to ensure the results were interpretable.

4 RESULTS

4.1 RQ1: Mean levels of experiences and their sources of variation and correlations

In model 1, we examined students' experiences: their mean levels, sources of variability and correlations at the student, survey, and residual levels. Because we were interested in part in the mean levels of the three variables for students' momentary experiences (their frustration, self-efficacy, and situational interest), for this model we did not standardize the variables as we did for the subsequent models. These estimates are shown in Table 1 (B parameters represent means, SD parameters standard deviations, and r parameters are correlations).

Table 1: Model 1: ESM experiences - Means, SDs, and correlations

Parameter	Model 1 (Experiences)
B - Frustration	2.668* (2.401 - 2.937)
B - Self-Efficacy	4.089* (3.878 - 4.297)
B - Situational Interest	4.238* (4.038 - 4.433)
SD - Residual: Frustration	0.573* (0.547 - 0.602)
SD - Residual: Self-Efficacy	0.584* (0.555 - 0.613)
SD - Residual: Situational Interest	0.514* (0.49 - 0.54)
SD - Person: Frustration	1.087 (0.918 - 1.295)
SD - Person: Self-Efficacy	0.855 (0.719 - 1.009)
SD - Person: Situational Interest	0.831 (0.703 - 0.99)
SD - Signal: Frustration	0.319 (0.22 - 0.46)
SD - Signal: Self-Efficacy	0.196 (0.133 - 0.283)
SD - Signal: Situational Interest	0.163 (0.113 - 0.233)
r - Person: Frustration - Self-Efficacy	-0.617* (-0.748 - -0.455)
r - Person: Frustration - Situational Interest	-0.365* (-0.552 - -0.137)
r - Person: Self-Efficacy - Situational Interest	0.665* (0.506 - 0.784)
r - Signal: Frustration - Self-Efficacy	-0.888* (-0.99 - -0.66)
r - Signal: Frustration - Situational Interest	-0.893* (-0.992 - -0.653)
r - Signal: Self-Efficacy - Situational Interest	0.944* (0.792 - 0.997)
r - Residual: Frustration-Self-Efficacy	-0.248* (-0.315 - -0.18)
r - Residual: Frustration-Situational Interest	-0.095* (-0.166 - -0.025)
r - Residual: Self-Efficacy-Situational Interest	0.377* (0.312 - 0.437)

4.1.1 Mean levels of students' experiences. The mean levels (see Table 1) of these variables indicate differences between the three experiences. We note that this and the following three tables report estimates with 95% credible intervals. Asterisks indicate that the effect is statistically significant based upon the ROPE hypothesis testing procedure. From this table, we can see that students' self-reported frustration (measured on a one to five scale) was equal to 2.668, whereas their self-efficacy and situational interest were each above 4. Thus, overall, students reported a moderate degree of frustration, and a high degree of self-efficacy and situational interest.

4.1.2 Sources of variation and correlations for measures of students' experiences. In addition to these mean levels, we emphasize two other types of statistics at this stage of reporting the results: the *SD* of the estimates for students' frustration, self-efficacy, and situational interest at the student, signal, and residual levels, as well as the correlation among the residuals for these three experiences at those same levels. In this way, these statistics illustrate an affordance of ESM (and other intensive longitudinal data): the ability to understand what about students' experiences can be attributed to students as individuals, what can be attributed to the particular moment at which students were signaled (a moment often following the submission of key assignment), and how much unexplained (or, residual) variance remains after accounting for these factors.

We found that there was substantial variation at the student level; on average, the student-level estimates (for students' mean levels of the variables for the three experiences) varied with a *SD* between .8 (for self-efficacy and situational interest) and 1.0 (for frustration), indicating that students' average levels of experiences vary substantially. There was less variation at the signal level, with unexplained variability at a moderate level. From these estimates, we can calculate an *Intra-Class Correlation*, or ICC, which represents the proportion of variance in students' experiences at the different levels included in the analysis. Around 50% of the variability in students' experiences was attributed to the person-level; around 10% at the signal level; and, thus, around 40% at the residual level.

We also inspected how these variables were correlated. At the person level, students who experienced more frustration (over the entire semester) also reported lower self-efficacy (with a moderate-strong correlation of -0.617); likewise when students reported a high-degree of self-efficacy, they were also likely to report a high degree of situational interest, (with a moderate-strong correlation of 0.665). At the signal level, the three experiences were strongly correlated: frustration and self-efficacy were negatively correlated, while self-efficacy and situational interest and self-efficacy and situational interest were both strongly positively correlated.

4.2 RQ1: Relations with experiences

Having investigated the mean levels of students' momentary experiences and their sources of variation and correlations, we next incorporated other variables to predict these experiences. We examined predictors of students' frustration, self-efficacy, and situational interest in two models. The first (model 2) added only students' self-reported gender and a dichotomous variable for their course to model 1: thus, these models can inform us about the extent to which there are gender- and course-related differences in students' experiences. We describe the results from this model first. We note that all continuous variables in these models were standardized to facilitate the interpretation of the effects relative to one another. These results are shown in Table 2.

Table 2: Model 2 and 3: Regression model parameters - ESM outcomes

Outcome	Parameter	Model 2 (Experiences)	Model 3 (Experiences)
Frustration	Intercept	0.339 (-0.043 - 0.719)	0.067 (-0.332 - 0.505)
Frustration	Gender - Male	-0.414 (-0.845 - 0.022)	-0.056 (-0.538 - 0.42)
Frustration	Course - CS-B	-0.075 (-0.79 - 0.622)	0.039 (-0.626 - 0.693)
Frustration	Initial Self-Efficacy		-0.297* (-0.517 - -0.066)
Frustration	Initial Interest		-0.125 (-0.325 - 0.086)
Self-Efficacy	Intercept	-0.443* (-0.817 - -0.084)	-0.09 (-0.435 - 0.265)
Self-Efficacy	Gender - Male	0.573* (0.194 - 0.988)	0.116 (-0.273 - 0.511)
Self-Efficacy	Course - CS-B	-0.121 (-0.736 - 0.581)	-0.252 (-0.804 - 0.273)
Self-Efficacy	Initial Self-Efficacy		0.305* (0.116 - 0.481)
Self-Efficacy	Initial Interest		0.334* (0.163 - 0.502)
Situational Interest	Intercept	-0.21 (-0.606 - 0.195)	-0.055 (-0.457 - 0.373)
Situational Interest	Gender - Male	0.253 (-0.2 - 0.696)	0.047 (-0.428 - 0.533)
Situational Interest	Course - CS-B	-0.205 (-0.931 - 0.483)	-0.216 (-0.858 - 0.423)
Situational Interest	Initial Self-Efficacy		-0.017 (-0.245 - 0.202)
Situational Interest	Initial Interest		0.459* (0.245 - 0.664)

4.2.1 Course and gender associations with students' experiences. As reported in Table 2, we found that there were statistically significant gender differences in students' self-efficacy outcome. Particularly, throughout the semester, students who reported their gender as male had self-efficacy 0.573 *SD* higher than those who reported their gender as female. Male students also reported -0.414 *SD* lower frustration (though the credible interval for this effect ranged from -0.845 - 0.022). There were not statistically significant differences between CS1A and CS1B. We do not report the effects for the parameters reported above (for the mean levels of students' experiences and their sources of variation and correlations) as these did not differ in terms of the magnitude or the statistical significance of the effects from those reported in Table 1.

4.2.2 The effects of students' initial self-efficacy and initial interest. In model 3 (also shown in Table 2), we added parameters for two students' initial self-efficacy and interest as predictors of the three momentary experiences. Thus, this model accounts for how the degree to which students began the class with higher (or lower) self-efficacy and interest (measured via a beginning-of-class survey) affects their in-class experiences of frustration, self-efficacy, and situational interest (measured via ESM). These models showed that students' initial self-efficacy was negatively associated with their experiences of frustration: For every one *SD* increase in students' initial self-efficacy, students reported -0.297 *SD* less frustration during the semester. Moreover, for every one *SD* increase in students' initial self-efficacy, students reported 0.305 *SD* greater self-efficacy throughout the semester. Students' initial interest was positively associated with their self-efficacy (β = 0.334 *SD*) and situational interest (β = 0.459 *SD*) during the semester. Thus, these models show that students do not have the same in-class experiences: Students' initial self-efficacy and interest-related beliefs work to shape the way they experience the course.

Another aspect of the findings from model 3 bear mention. In particular, the effect of students' gender upon their self-efficacy experiences decreased from 0.573 to 0.116 - an approximately five-fold decrease in the magnitude of the gender effect. Moreover, this effect was no longer statistically significant. Though the credible interval for the gender effect of students' sense of frustration was wide, the magnitude of the relationship decreased (once students' initial self-efficacy and interest were added to the model) from -0.414 to -0.056, an approximately seven-fold decrease; and the gender effect on students' situational interest decreased from 0.253 to 0.046. In short, the results in model 3 demonstrate that what initially (in model 2) appeared to be gender-related differences in students' experiences are due largely to initial self-efficacy and interest-related differences.

4.2.3 The proportion of variance explained in students' experiences. To understand the proportion of variance explained by these models, we calculated the R^2 values for the three outcomes—frustration, self-efficacy, and situational interest—for model 1. The R^2 values, which include both *fixed effects* (those that are estimated to be the same across the entire data set) and *random effects* (those that do not vary across the person or signal levels, or are at the residual level), were high, ranging from .655 for frustation (indicating that around 65% of the variation in frustration was explained by the model) to 0.733 for situational interest (See Appendix A). These values—referred to as the total R^2 values—were comparable for models 2 and 3 (see Appendix A). In all, these suggest that a substantial proportion of the variance in students' outcomes can be explained by knowing the student and signal.

In this way, these R^2 values represent the maximum variation that can be explained, knowing the person (student) and signal identifier. In addition to knowing how much variation was explained overall, we were also interested in the proportion of the variance explained by the fixed effects—those for students' self-reported gender, the course indicator, and students' initial interest and self-efficacy. We focused on these *marginal* R^2 values for model 3, which included these four fixed effects. These marginal R^2 values ranged from 0.156 (for frustration) to 0.318 (for self-efficacy). While smaller than the total R^2 values, which include the random effects for the person and signal, these suggest that a still sizable proportion of the variation in students' experiences can be explained by the other constructs, including students' characteristics and inclinations at the outset of the class.

4.3 RQ2: Associations between student characteristics and end-of-course outcomes

Having examined students' momentary experiences, both descriptively and as they relate to predictors, we next sought to understand how these experiences related to key end-of-course outcomes: students' interest in CS following the semester and their final score in their course. We present these analyses in two tables.

In table 3, we report estimates of the relationships between the student level fixed effects and the two end-of-course outcomes: students' interest in CS and their final course score. Like for models 2 and 3, all of the continuous variables were standardized, to facilitate interpretation of the effects.

These estimates show a similar pattern to those in model 3: while gender effects were minimal (and not statistically significant), there were strong relations between students' initial self-efficacy and interest, and their end-of course outcomes. Particularly, students with higher initial self-efficacy reported higher end-of-course interest (accounting for their initial interest; β = 0.356 *SD*). Students with higher initial interest reported higher end-of-course interest (β = 0.195 *SD*) and higher final course scores (β = 0.459 *SD*)). The negative relationship between initial self-efficacy and course grade is harder to explain. In short, students' initial characteristics relate to two end-of-course outcomes.

4.4 RQ2: Relations of students' experiences with their end-of-course outcomes

While notable, the findings just described do not provide an answer to the core question we asked: How do students' experiences within-class—which CS instructors have more of an opportunity to influence—relate to these outcomes? In Table 4, we present the results for the residual correlations between students' frustration, self-efficacy, and situational interest (measured via ESM) with the two end-of-course outcomes. These depict the extent to which students with higher (or lower) levels of frustrating, self-efficacy supporting (or harming), and interesting experiences relate to each of the two outcomes (in separate models).

Model 4 shows that students who reported higher self-efficacy during the semester reported higher end-of-course interest in CS (r = 0.278) and that students who reported experiencing greater situational interest reported much higher end-of-course interest (r = 0.607). Students who experienced more frustration were much more likely to achieve a lower final course score (r = -0.469). Notably, these models included the fixed effects and the mean levels of students' experiences and their sources of variation and correlations; these are not reported because—like for the results reported in models 2 and 3—the magnitude and statistical significance of these effects did not differ from those reported in the above models. Substantively, this indicates that students' experiences in-class impact their end-of-course outcomes above and beyond the effect of the characteristics and inclinations they bring with them at the outset of class. We discuss the implications of these and other results next.

4.4.1 The proportion of variance explained in end-of-course outcomes. Like for students' experiences, we calculated the proportion of variance explained in end-of-course outcomes (See Appendix A). We only report the total R^2 values, as these are the same as the marginal R^2 values because these outcomes were neither grouped within persons (students) or signals. For students' end-of-course interest, 0.241 (or, around 24.1%) of the variance was explained by the fixed effects; for students' final score, the value was 0.162.

5 DISCUSSION

This study was predicated on the lack of prior research on the momentary experiences students have in CS1 classes. Unlike research examining, for instance, changes in interest between the beginning and end of a class, or students' retrospective reports having particular types of experiences over the course of the semester, the present study used ESM as a method to understand students' experiences in the moment. This study treated students' granular experiences of frustration, self-efficacy, and interest as outcomes of importance on their own. We showed that CS1 students, generally, reported high levels of interest and self-efficacy, and moderate levels of frustration. Moreover, we showed that there is substantial person-to-person variability in these experiences, and that students who tend to have more frustrating experiences over the semester report lower self-efficacy and that students who report higher self-efficacy report higher interest.

We not only modeled these experiences of frustration, self-efficacy, and interest as a means of describing them (and their relations with one another), but also as an outcome that may vary based upon students' self-reported gender and their initial self-efficacy and interest. Critically, we believe, we showed that differences in these experiences that may be attributed to students' gender greatly diminish when also considering differences in students' initial levels of self-efficacy and interest. Finally, we connected these experiences that students reported having to two course-long outcomes, one for students' end-of-course interest, which we considered to be a measure of their individual interest that may help us to understand how students may continue to engage with CS in the following semester or year, as well as students' final course score. We found that students who reported having more positive self-efficacy related experiences (even holding their beginning of the course self-efficacy constant) reported a higher degree of interest in CS at the end of the semester. We also found that students whose interest was higher during the semester (accounting for their initial interest) were much more likely to be interested in CS at the conclusion of the semester. Moreover, students who reported less frustration were, generally, higher-achieving in terms of the course outcome.

5.1 Key Findings

5.1.1 The importance of the experiences students have in CS1. One primary takeaway from the results just presented, is that students' momentary experiences are related to their interest in computing, when controlling for their initial levels of interest. Model 4, particularly, showed that there were significant effects of initial self-efficacy, initial interest, and course on students' interest in computing at the end of the course. This is accounting for the random effects between students and between ESM survey occasions.

Table 3: Model 4 and 5: Fixed effects - end-of-course outcomes

Parameter	Model 4 (Interest)	Model 5 (Course Score)
Intercept	0.01 (-0.11 - 0.138)	0.133 (0.006 - 0.265)
Gender - Male	0.01 (-0.14 - 0.152)	-0.197* (-0.363 - -0.05)
Course - CS-B	-0.315* (-0.559 - -0.084)	0.261 (0.002 - 0.49)
Initial Self-Efficacy	0.356* (0.287 - 0.428)	-0.184* (-0.251 - -0.108)
Initial Interest	0.195* (0.123 - 0.262)	0.459* (0.384 - 0.536)

Table 4: Model 4 and 5: Residual correlations - end-of-course outcomes

Parameter	Model 4 (Interest)	Model 5 (Course Score)
SD - End-of-Course Outcome	0.851* (0.809 - 0.891)	0.921* (0.876 - 0.965)
r - Frustration - End-of-Course Outcome	-0.188 (-0.449 - 0.097)	-0.469* (-0.629 - -0.265)
r - Self-Efficacy - End-of-Course Outcome	0.278* (0.056 - 0.474)	0.205 (-0.034 - 0.413)
r - Situational Interest - End-of-Course Outcome	0.607* (0.444 - 0.731)	0.211 (-0.108 - 0.476)

Beyond this, the residual correlation values tell us that students' momentary experiences— particularly their momentary experiences of self-efficacy and situational interest—are significantly related to their interest in computing, accounting for the other fixed and random effects. Thus, while we knew prior to this study that initial self-efficacy and interest would impact the interest at the end of the course given prior research [42, 86], we did not know about these effects of momentary experiences of self-efficacy and situational interest that exist even accounting for the initial levels. This is a promising finding because it gives empirical credence to the notion that self-efficacy fluctuates and responds to context in the course. Moreover, this is promising because it is at the level of these momentary experiences that CS instructors have the opportunity to influence students' experiences, as opposed to the self-efficacy they bring to the class with them. The connection between students' experiences of situational interest with their end of course interest points towards the way that aggregate experiences of situational interest can build up towards emerging individual interest, or transitioning from phase 2 to phase 3 in Hidi and Renninger's four-phase model [41].

In addition to findings related to students' end-of-course interest, the model for students' achievement (model 5) showed the significant effects of these factors on the end of course grade outcome. Contrasting with model 4, the significant fixed effects were for gender but not for course, as well as initial self-efficacy and interest. Above and beyond these fixed effects (and the random effects for student and survey occasion), the residual correlations tell us the effects of students' momentary experiences on the course grade outcome. Contrary to model 4, there was not a significant residual effect of momentary self-efficacy or situational interest on course grade, but there was a significant residual effect of frustration. As with self-efficacy, momentary frustration is a dimension of students' experiences that instructors have the opportunity to influence, so it is a positive finding to see this strong effect of in-class experiences, even when accounting for baseline student factors from the beginning of the semester.

The differences between the two end-of-course outcome models point to the intuitive conclusion that leaving a CS1 course with a good grade is not the same as leaving the course with an interest in the field. Different factors, we see from the results of our models, play into these two outcomes. Taken together though, these two outcome models point to the importance of all three dimensions of students' momentary experiences that we examined. Momentary self-efficacy, situational interest, and experiences of frustration all significantly relate to one important outcome. Thus all three dimensions are ones that we need to pay attention to in computing education research and pedagogy.

5.2 Self-efficacy and interest gaps, rather than gender gaps

In addition to the importance of students' experiences in CS1, another key takeaway from this study is that the gender differences observed in both the affective experiences, and with respect to the computing interest outcome at the end of the course, were substantially reduced when students' initial self-efficacy was included in the model. Initially, female students experienced greater momentary frustration and lower momentary self-efficacy across the ESM surveys, but when accounting for the initial self-efficacy levels these differences nearly disappeared, suggesting that what initially appeared to be gender differences are more accurately described as self-efficacy differences. This is obviously a better result for computing education because self-efficacy is something that can be addressed pedagogically.

These findings are constructive in that computing educators can address self-efficacy by addressing the way that self-efficacy is shaped throughout the semester by students' experiences in CS1. Our results suggest that gender differences in interest in computing can be addressed by supporting self-efficacy, and this need not be done by attempting to correct initial gender self-efficacy gaps in a CS1 course, which we know have been persistent in this field [7, 43] (although that may also be viable), but rather CS educators can better support women by building in more self-efficacy supporting experiences throughout the course. Such pedagogical approaches to supporting the goal of broadening participation would nicely complement the ongoing outreach efforts towards this goal.

5.3 Limitations and recommendations for future research

One limitation of this study is the sample, which is not a random sample from which we can straightforwardly make inferences to a national (much less an international) 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 CS1A and CS1B courses who agreed to take part in the study. That being the case, we cannot say with certainty whether the results produced with this subset of students would match what would be observed if the entirety of both CS1 classes had taken part in the study. That being said, given that this study is more of an exploratory investigation of students' CS1 experiences, rather than say, an evaluation of an intervention, it seems less likely that the self-selection bias would strongly change the conclusions drawn in this paper. There is no reason to believe that experiences of self-efficacy and frustration would not be related to end-of-course outcomes for students without the propensity to opt in to participate in our study; however, we consider this to be only a first study that establishes the findings we reported as conjectures. Future research is needed to determine whether the findings observed in this study would generalize to other contexts. Another potential limitation concerns our treatment of the data from our Likert-type ESM variables as approximations of continuous variables. While it is possible for this to introduce bias into the estimation, the distribution of the five-point variables makes it such that we think any bias resulting from this choice is likely to be minimal [1].

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 momentary experiences in CS1, but there are certainly many others that may be important that would be worth examining in future research. For example, we do not have measures of the context in these classes that might inform why we see the experiences that we did. This could involve ways of measuring 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' motivational outcomes in prior research. There are certainly other student factors that would be worth examining in future research as well, such as prior experience in CS, metacognitive self-regulation, goal orientation to name just a few. We hope that future research can continue to build a more nuanced picture of students' momentary experiences in CS1, with an eye towards generating pedagogical concepts and resources to make these classes as supportive of students as possible.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 1937700.

REFERENCES

[1] 2020. Can an Ordinal Likert Scale be a Continuous Variable? https://www.statisticssolutions.com/can-an-ordinal-likert-scale-be-a-continuous-variable/
[2] Karen Adams. 2004. Modelling success: enhancing international postgraduate research students' self-efficacy for research seminar presentations. *Higher Education Research & Development* 23, 2 (2004), 115–130. https://doi.org/10.1080/0729436042000206618
[3] Mary Ainley. 2007. Being and feeling interested: Transient state, mood, and disposition. In *Emotion in Education*, Paul A Schultz and Reinhard Pekrun (Eds.). 141–157.
[4] Mary Ainley and John Ainley. 2011. Student engagement with science in early adolescence: The contribution of enjoyment to students' continuing interest in learning about science. *Contemporary Educational Psychology* 36, 1 (2011), 4–12. https://doi.org/10.1016/j.cedpsych.2010.08.001
[5] Heidi L. Andrade. 2019. A Critical Review of Research on Student Self-Assessment. *Frontiers in Education* 4, August (2019), 1–13. https://doi.org/10.3389/feduc.2019.00087
[6] 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
[7] 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
[8] 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
[9] Jo Boaler. 2013. Ability and Mathematics: the mindset revolution that is reshaping education. *Forum* 55, 1 (2013), 143. https://doi.org/10.2304/forum.2013.55.1.143
[10] Nigel Bosch and Sidney D'Mello. 2013. Sequential patterns of affective states of novice programmers. *CEUR Workshop Proceedings* 1009, Aiedcs (2013), 1–10.
[11] Nigel Bosch and Sidney D'Mello. 2017. The Affective Experience of Novice Computer Programmers. *International Journal of Artificial Intelligence in Education* 27, 1 (2017), 181–206. https://doi.org/10.1007/s40593-015-0069-5
[12] Paul-Christian Bürkner. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software* 80, 1 (2017), 1–28.
[13] James Carifio and Rocco Perla. 2008. Resolving the 50-year debate around using and misusing Likert scales. *Medical Education* 42, 12 (2008), 1150–1152. https://doi.org/10.1111/j.1365-2923.2008.03172.x
[14] Lori Carter. 2007. Why students with an apparent aptitude for computer science don't choose to major in computer science. *Proceedings of the Thirty-Seventh SIGCSE Technical Symposium on Computer Science Education* (2007), 27–31. https://doi.org/10.1145/1121341.1121352
[15] Simon Cassidy. 2015. Resilience building in students: The role of academic self-efficacy. *Frontiers in Psychology* 6, NOV (2015), 1–14. https://doi.org/10.3389/fpsyg.2015.01781
[16] Sapna Cheryan, Allison Master, and Andrew N. Meltzoff. 2015. Cultural stereotypes as gatekeepers: Increasing girls' interest in computer science and engineering by diversifying stereotypes. *Frontiers in Psychology* 6, FEB (2015), 1–8. https://doi.org/10.3389/fpsyg.2015.00049
[17] Sapna Cheryan, Victoria C Plaut, Paul G Davies, and Claude M Steele. 2009. Ambient belonging: how stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology* 97, 6 (2009), 1045–1060. https://doi.org/10.1037/a0016239
[18] Anne Marie M. Conley. 2012. Patterns of motivation beliefs: Combining achievement goal and expectancy-value perspectives. *Journal of Educational Psychology* 104, 1 (2012), 32–47. https://doi.org/10.1037/a0026042
[19] 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
[20] Mihaly Csikszentmihalyi and Reed Larson. 2014. Validity and Reliability of the Experience-Sampling Method. In *Flow and the Foundations of Positive Psychology.* 1–298. https://doi.org/10.1007/978-94-017-9088-8
[21] 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.
[22] Ivana Di Leo, Krista R. Muis, Cara A. Singh, and Cynthia Psaradellis. 2019. Curiosity… Confusion? Frustration! The role and sequencing of emotions during mathematics problem solving. *Contemporary Educational Psychology* 58, March (2019), 121–137. https://doi.org/10.1016/j.cedpsych.2019.03.001
[23] 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.
[24] Remy Dou, Karina Bhutta, Monique Ross, Laird Kramer, and Vishodana Thamotharan. 2020. The Effects of Computer Science Stereotypes and Interest on Middle School Boys' Career Intentions. *ACM Transactions on Computing Education* 20, 3 (2020). https://doi.org/10.1145/3394964
[25] 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

[26] 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.

[27] Andrea Forte and Mark Guzdial. 2005. Motivation and nonmajors in computer science: Identifying discrete audiences for introductory courses. *IEEE Transactions on Education* 48, 2 (2005), 248–253. https://doi.org/10.1109/TE.2004.842924

[28] Brian M. Galla, Jamie Amemiya, and Ming Te Wang. 2018. Using expectancy-value theory to understand academic self-control. *Learning and Instruction* 58, May (2018), 22–33. https://doi.org/10.1016/j.learninstruc.2018.04.004

[29] 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

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

[31] Thomas Goetz, Anne C. Frenzel, Nathan C. Hall, and Reinhard Pekrun. 2008. Antecedents of academic emotions: Testing the internal/external frame of reference model for academic enjoyment. *Contemporary Educational Psychology* 33, 1 (2008), 9–33. https://doi.org/10.1016/j.cedpsych.2006.12.002

[32] Thomas Goetz, Anne C. Frenzel, Heidrun Stoeger, and Nathan C. Hall. 2010. Antecedents of everyday positive emotions: An experience sampling analysis. *Motivation and Emotion* 34, 1 (2010), 49–62. https://doi.org/10.1007/s11031-009-9152-2

[33] Google. 2014. Women Who Choose Computer Science — What Really Matters (The Critical Role and Exposure).

[34] Jamie Gorson and Eleanor O'Rourke. 2020. Why do CS1 Students Think They're Bad at Programming?: Investigating Self-efficacy and Self-assessments at Three Universities. *ICER 2020 - Proceedings of the 2020 ACM Conference on International Computing Education Research* (2020), 170–181. https://doi.org/10.1145/3372782.3406273

[35] Jamie Gorson and Eleanor O'Rourke. 2021. CS1 Student Assessments of Themselves Relative to Others: The Role of Self-Critical Bias, Gender, and Sense of Belonging. In *International Society of the Learning Sciences*.

[36] Joseph F. Grafsgaard, Joseph B. Wiggins, Kristy Elizabeth Boyer, Eric N. Wiebe, and James C. Lester. 2013. Automatically recognizing facial indicators of frustration: A learning-centric analysis. *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013* (2013), 159–165. https://doi.org/10.1109/ACII.2013.33

[37] Daniel Graziotin, Xiaofeng Wang, and Pekka Abrahamsson. 2014. Pattern detection for conceptual schema recovery in data-intensive systems. *Journal of Software: Evolution and Process* 26, 12 (2014), 1172–1192. https://doi.org/10.1002/smr

[38] Mark Guzdial. 2013. Exploring hypotheses about media comptation. In *ICER 2013 - Proceedings of the 2013 ACM Conference on International Computing Education Research*. 19–26. https://doi.org/10.1145/2493394.2493397

[39] Markku S Hannula. 2015. Emotions in Problem Solving. In *Selected Regular Lectures from the 12th International Congress on Mathematical Education*, Sun Joo Cho (Ed.). Springer International Publishing, 269–288. https://doi.org/10.1007/978-3-319-17187-6

[40] Joel M. Hektner, Jennifer A. Schmidt, and Mihaly Csikszentmihalyi. 2007. *Experience Sampling Method: Measuring the quality of everyday life*. Sage.

[41] Suzanne Hidi and Ann Renninger. 2006. The Four-Phase Model of Interest Development. *Educational Psychologist* 41, 2 (2006), 111–127. https://doi.org/10.1207/s15326985ep4102 arXiv:0710.4428v1

[42] 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

[43] 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]

[44] 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

[45] Martin H. Jones, Shannon R. Audley-Piotrowski, and Sarah M. Kiefer. 2012. Relationships among adolescents' perceptions of friends' behaviors, academic self-concept, and math performance. *Journal of Educational Psychology* 104, 1 (2012), 19–31. https://doi.org/10.1037/a0025596

[46] Hayal Z. Kackar, Lee Shumow, Jennifer A. Schmidt, and Janel Grzetich. 2011. Age and gender differences in adolescents' homework experiences. *Journal of Applied Developmental Psychology* 32, 2 (2011), 70–77. https://doi.org/10.1016/j.appdev.2010.12.005

[47] Y. Kafai, D. Fields, and K. Searle. 2014. Electronic textiles as disruptive maker activities in schools. *Harvard Educational Review* 84, 4 (2014), 532–557.

[48] Yasmin B. Kafai, Eunkyoung Lee, Kristin Searle, Deborah Fields, Eliot Kaplan, and Debora Lui. 2014. A crafts-oriented approach to computing in high school: Introducing computational concepts, practices, and perspectives with electronic textiles. *ACM Transactions on Computing Education* 14, 1 (2014), 1–20. https://doi.org/10.1145/2576874

[49] Sandra Katz, David Allbritton, John Aronis, Christine Wilson, and Mary Lou Soffa. 2006. Gender, Achievement, and Persistence in an Undergraduate Computer Science Program. *Data Base for Advances in Information Systems* 37, 4 (2006), 42–57. https://doi.org/10.1145/1185335.1185344

[50] Jen Katz-Buonincontro and Rajashi Ghosh. 2014. Using workplace experiences for learning about affect and creative problem solving: Piloting a four-stage model for management education. *International Journal of Management Education* 12, 2 (2014), 127–141. https://doi.org/10.1016/j.ijme.2014.03.003

[51] Jen Katz-Buonincontro and Joel M. Hektner. 2014. Using experience sampling methodology to understand how educational leadership students solve problems on the fly. *Journal of Educational Administration* 52, 3 (2014), 379–403. https://doi.org/10.1108/JEA-12-2012-0135

[52] 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

[53] 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

[54] 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

[55] John Kruschke. 2014. Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan. (2014).

[56] 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

[57] Woogul Lee, Myung Jin Lee, and Mimi Bong. 2014. Testing interest and self-efficacy as predictors of academic self-regulation and achievement. *Contemporary Educational Psychology* 39, 2 (2014), 86–99. https://doi.org/10.1016/j.cedpsych.2014.02.002

[58] 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

[59] 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.

[60] 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

[61] Lisa Linnenbrink-Garcia and Reinhard Pekrun. 2011. Students' emotions and academic engagement: Introduction to the special issue. *Contemporary Educational Psychology* 36, 1 (2011), 1–3. https://doi.org/10.1016/j.cedpsych.2010.11.004

[62] Alex Lishinski, Joshua M. Rosenberg, Michael Mann, Omiya Sultana, and Joshua Dunn. 2021. How CS1 Students Experienced COVID-19 in the Moment. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*. 1254–1254.

[63] Alex Lishinski and Aman Yadav. 2018. Motivation, Attitudes, and Dispositions. In *Handbook of Computing Education Research*, Sally Fincher and Anthony Robins (Eds.). Cambridge University Press.

[64] Alex Lishinski, Aman Yadav, and Richard Enbody. 2017. Students' Emotional Reactions to Programming Projects in Introduction to Programming: Measurement Approach and Influence on Learning Outcomes. *Proceedings of the 2017 ACM Conference on International Computing Education Research* (2017), 30–38.

[65] 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).

[66] 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

[67] Jane Margolis, Allan Fisher, and Faye Miller. 2000. The Anatomy of Interest : Women in Undergraduate Computer Science Author (s): Jane Margolis , Allan Fisher and Faye Miller Source : Women ' s Studies Quarterly , Vol . 28 , No . 1 / 2 , Building Inclusive Science (Spring - Published by : The Feminist P. *Women's Studies Quarterly* 28, 1/2 (2000), 104–127.

[68] Allison Master, Sapna Cheryan, and Andrew N. Meltzoff. 2016. Computing whether she belongs: Stereotypes undermine girls' interest and sense of belonging in computer science. *Journal of Educational Psychology* 108, 3 (2016), 424–437. https://doi.org/10.1037/edu0000061

[69] T Mcklin, Lauren Mccall, Taneisha Lee, Brian Magerko, Michael Horn, and Jason Freeman. 2021. Leveraging Prior Computing and Music Experience for Situational Interest Formation. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*. 928–933.

[70] Carolina Milesi, Lara Perez-Felkner, Kevin Brown, and Barbara Schneider. 2017. Engagement, persistence, and gender in computer science: Results of a smartphone ESM study. *Frontiers in Psychology* 8, APR (2017), 1–9. https://doi.org/10.3389/fpsyg.2017.00602

[71] Krista R. Muis, Marianne Chevrier, and Cara A. Singh. 2018. The Role of Epistemic Emotions in Personal Epistemology and Self-Regulated Learning. *Educational Psychologist* 53, 3 (2018), 165–184. https://doi.org/10.1080/00461520.2017.1421465

[72] 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

[73] Steve Murphy, Amy MacDonald, Cen Audrey Wang, and Lena Danaia. 2019. Towards an Understanding of STEM Engagement: a Review of the Literature on Motivation and Academic Emotions. *Canadian Journal of Science, Mathematics and Technology Education* 19, 3 (2019), 304–320. https://doi.org/10.1007/s42330-019-00054-w

[74] Diane Nahl. 2005. Affective Load. In *Theories of information behavior.*, Karen E. Fisher, Sanda Erdelez, and Lynne McKechnie (Eds.). Information Today, Medford, N.J., 39–43.

[75] 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

[76] Geoff Norman. 2010. Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education* 15, 5 (2010), 625–632. https://doi.org/10.1007/s10459-010-9222-y

[77] 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

[78] Reinhard Pekrun, Aisling Cusack, Kou Murayama, Andrew J. Elliot, and Kevin Thomas. 2014. The power of anticipated feedback: Effects on students' achievement goals and achievement emotions. *Learning and Instruction* 29 (2014), 115–124. https://doi.org/10.1016/j.learninstruc.2013.09.002

[79] Reinhard Pekrun, Stephanie Lichtenfeld, Herbert W. Marsh, Kou Murayama, and Thomas Goetz. 2017. Achievement Emotions and Academic Performance: Longitudinal Models of Reciprocal Effects. *Child Development* 88, 5 (2017), 1653–1670. https://doi.org/10.1111/cdev.12704

[80] Michael Pelch. 2018. Gendered differences in academic emotions and their implications for student success in STEM. *International Journal of STEM Education* 5, 1 (2018). https://doi.org/10.1186/s40594-018-0130-7

[81] Markeya S. Peteranetz, Shiyuan Wang, Duane F. Shell, Abraham E. Flanigan, and Leen Kiat Soh. 2018. Examining the impact of computational creativity exercises on college computer science students' learning, achievement, self-efficacy, and creativity. *SIGCSE 2018 - Proceedings of the 49th ACM Technical Symposium on Computer Science Education* 2018-Janua (2018), 155–160. https://doi.org/10.1145/3159450.3159459

[82] 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

[83] 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

[84] 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

[85] K. Ann Renninger and Suzanne Hidi. 2011. Revisiting the conceptualization, measurement, and generation of interest. *Educational Psychologist* 46, 3 (2011), 168–184. https://doi.org/10.1080/00461520.2011.587723

[86] K. Ann Renninger and Suzanne E. Hidi. 2020. To Level the Playing Field, Develop Interest. *Policy Insights from the Behavioral and Brain Sciences* 7, 1 (2020), 10–18. https://doi.org/10.1177/2372732219864705

[87] 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

[88] Samara Ruiz, Sven Charleer, Maite Urretavizcaya, Joris Klerkx, Isabel Fernández-Castro, and Erik Duval. 2016. Supporting learning by considering emotions. (2016), 254–263. https://doi.org/10.1145/2883851.2883888

[89] Ulrich Schiefele. 2001. The Role of Interest in Motivation and Learning. In *Intelligence and personality: Bridging the gap in theory and measurement*, J.M. Collis and S. Messick (Eds.). Erlbaum, Mahwah, NJ, 163–194.

[90] 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.

[91] Dale H. Schunk and Frank Pajares. 2002. The Development of Academic Self-Efficacy. In *Development of Achievement Motivation*. 15–31. https://doi.org/10.12930/0271-9517-8.1.11

[92] Amy Schweinle, Julianne Turner, and Debra Meyer. 2008. Understanding young adolescents' optimal experiences in academic settings. *Journal of Experimental Education* 77, 2 (2008), 125–146. https://doi.org/10.3200/JEXE.77.2.125-146

[93] Christie Scollon, Chu Kim-Prieto, and Ed Diener. 2003. Experience Sampling: Promises and Pitfalls, Strengths and Weaknesses. *Journal of Happiness Studies* 4, 1 (2003), 5–34. http://ideas.repec.org/a/spr/jhappi/v4y2003i1p5-34.html

[94] Duane F Shell, Abraham E Flanigan, and Markeya S Peteranetz. 2016. Students ' Initial Course Motivation and Their Achievement and Retention in College CS1 Courses. *Proceedings of the 47th ACM Technical Symposium on Computer Science Education (SIGCSE '16)* (2016), 639–644.

[95] Lee Shumow and Jennifer A Schmidt. 2008. Reading In Class & Out of Class: An Experience Sampling Method Study. *Middle Grades Research Journal* 3, 3 (2008), 97–120. http://proxy.geneseo.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true{&}db=eric{&}AN=EJ832302{&}site=ehost-live{&}scope=site

[96] 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

[97] Mohamad B. Sukariyah and Ghada Assaad. 2015. The Effect of Attribution Retraining on the Academic Achievement of High School Students in Mathematics. *Procedia - Social and Behavioral Sciences* 177, July 2014 (2015), 345–351. https://doi.org/10.1016/j.sbspro.2015.02.356

[98] Gail M. Sullivan and Anthony R. Artino. 2013. Analyzing and Interpreting Data From Likert-Type Scales. *Journal of Graduate Medical Education* 5, 4 (2013), 541–542. https://doi.org/10.4300/jgme-5-4-18

[99] 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 (2004), 111–133. https://doi.org/10.1207/s15326985ep3902

[100] 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

[101] 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.

[102] 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

[103] 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

[104] Kimberly Michelle Ying, Fernando J Rodríguez, Alexandra Lauren Dibble, Alexia Charis Martin, Kristy Elizabeth Boyer, Sanethia V Thomas, and Juan E Gilbert. 2021. Confidence, Connection, and Comfort: Reports from an All-Women's CS1 Class. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*. 699–705.

[105] 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

[106] Sabrina Zirkel, Julie A. Garcia, and Mary C. Murphy. 2015. Experience-Sampling Research Methods and Their Potential for Education Research. *Educational Researcher* 44, 1 (2015), 7–16. https://doi.org/10.3102/0013189X14566879

A VARIANCE EXPLAINED BY MODEL

Model	Frust. (Tot)	Frust. (Marg)	Self-Eff (Tot)	Self-Eff (Marg)	Sit. Int. (Total)	Sit. Int. (Marg)	Course Out.
1	0.67	0.00	0.66	0.00	0.73	0.00	
2	0.68	0.04	0.65	0.08	0.73	0.03	
3	0.68	0.16	0.65	0.32	0.73	0.23	
4	0.67	0.18	0.62	0.34	0.52	0.21	0.24
5	0.58	0.17	0.64	0.34	0.72	0.23	0.16