Using Electrodermal Activity Measurements to Understand Student Emotions While Programming

Jamie Gorson
Northwestern University
Evanston, Illinois, USA
jgorson@u.northwestern.edu

Kathryn Cunningham
Northwestern University
Evanston, Illinois, USA
kcunningham@northwestern.edu

Marcelo Worsley
Northwestern University
Evanston, Illinois, USA
marcelo.worsley@northwestern.edu

Eleanor O'Rourke
Northwestern University
Evanston, Illinois, USA
eorourke@northwestern.edu

ABSTRACT

Programming can be an emotional experience, particularly for undergraduate students who are new to computer science. While researchers have interviewed novice programmers about their emotional experiences, it can be difficult to pinpoint the specific emotions that occur during a programming session. In this paper, we argue that electrodermal activity (EDA) sensors, which measure the physiological changes that are indicative of an emotional reaction, can provide a valuable new data source to help study student experiences. We conducted a study with 14 undergraduate students in which we collected EDA data while they worked on a programming problem. This data was then used to cue the participants' recollections of their emotions during a retrospective interview about the programming experience. Using this methodology, we identified 21 distinct events that triggered student emotions, such as feeling anxiety due to a lack of perceived progress on the problem. We also identified common patterns in EDA data across multiple participants, such as a drop in their physiological reaction after developing a plan, corresponding with a calmer emotional state. These findings provide new information about how students experience programming that can inform research and practice, and also contribute initial evidence of the value of EDA data in supporting studies of emotions while programming.

CCS CONCEPTS

• **Social and professional topics → Computing education**.

KEYWORDS

introductory programming, electrodermal activity, emotion, frustration, retrospective interview

ICER 2022, August 7–11, 2022, Lugano and Virtual Event, Switzerland

ACM ISBN 978-1-4503-9194-8/22/08…$15.00
https://doi.org/10.1145/3501385.3543981

ACM Reference Format:
Jamie Gorson, Kathryn Cunningham, Marcelo Worsley, and Eleanor O'Rourke. 2022. Using Electrodermal Activity Measurements to Understand Student Emotions While Programming. In *Proceedings of the 2022 ACM Conference on International Computing Education Research V.1 (ICER 2022), August 7–11, 2022, Lugano and Virtual Event, Switzerland*. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3501385.3543981

1 INTRODUCTION

For introductory computer science students, code writing can produce a roller-coaster of emotions, from frustration and desperation to joy and pride [6, 7, 13, 16, 26]. Even though emotions are experienced in small moments, they have been shown to correlate with longer-term outcomes like project and course performance [6, 7, 25, 29], self-efficacy [27, 29], and self-assessed productivity [24]. For example, Lishinski et al. found that feelings of pride are positively correlated with project scores, while feelings of frustration are negatively correlated [29]. As a result, the emotions that students experience while programming are important for practitioners and computing education researchers to understand.

Identifying the specific causes of students' emotions while programming is key to improving their programming experiences. A number of studies have explored programmer emotions using a variety of methods, including asking general questions about the events that trigger emotions [22, 34], interrupting students while programming to learn about their emotions [20], and asking about affect during predetermined events [7]. While these studies provide a valuable foundation to our understanding of programmer emotions, they have a number of limitations. Participants often struggle to accurately recall their emotional reactions after the fact [36], but interrupting them as they work impacts the authenticity of the programming experience. However, we currently lack a method for analyzing emergent programmer emotions during authentic programming sessions with a moment-to-moment unit of analysis.

Measuring physiological reactions throughout a programming session may provide new insights into the emotional experiences of programmers. Emotions are not just cognitive reactions; they also create physiological changes in the body, like increases in heart rate and sweat production, which can be measured by sensors [9]. Physiological sensors provide continuous and fine-grained data that allow researchers to pinpoint particular moments when people experience emotions and analyze emotions across time. Since people may not always be aware of their emotions, physiological devices can measure emotional stimuli even when people are not cognizant of those experiences [5, 8].

With the increasing availability of sensors, cued-recall experts have suggested adding information provided by physiological data sources to retrospective interviews to assist in triggering student recall of events [4]. Physiological data provides an additional lens through which interviewers and participants can view participants' emotional experiences, which complements the cognitive and behavioral indicators that we can measure with surveys, interviews, log data, and observation [8, 9, 19]. Physiological data sources may help improve student recall in retrospective interviews by surfacing moments of high emotional responses for further discussion. However, we are not aware of any existing studies that have attempted to utilize physiological data sources with a cued-recall methodology in the computer science education domain.

In this paper, we use electrodermal activity (EDA) sensors to capture physiological data while students work on a programming problem. We then utilize this data to trigger student recall during a retrospective interview of the programming session. We use this methodology to answer two questions:

- RQ1: What events trigger students to experience emotions while programming?
- RQ2: How do students' remembered experiences align with visual inspection of EDA data?

Through our analysis, we identify 21 different events that trigger emotions as students program. We also demonstrate the relationship between EDA data and student experiences through broader patterns that emerged across participants. These findings have important implications for research and practice, and also suggest that EDA data can serve as a valuable resource for prompting student recollections during retrospective interviews.

2 BACKGROUND

Psychologists have studied emotions for decades, using tools like EDA, while computing education researchers have explored programmer emotions more recently. In this section, we briefly define emotions before discussing how emotions, and the triggers of emotions, have been studied with programmers. We then introduce EDA as a tool for measuring emotional responses, and finally describe how EDA data has been used to study programmers.

2.1 Definitions of emotion

While emotions are a common occurrence in everyday life, they are notoriously difficult to define scientifically [42]. There is agreement across theories that emotions involve multiple processes in the body, including both cognition and the work of the autonomous nervous system. This is reflected in the American Psychological Association's definition of an emotion, which is: "a complex reaction pattern, involving experiential, behavioral, and physiological elements, by which an individual attempts to deal with a personally significant matter or event" [1]. Similarly, Scherer's Component Process Model identifies five coordinated processes that make up emotions. The five processes are: cognitive appraisal of the situation, bodily symptoms, tendencies towards action, facial and vocal expression, and feelings [42]. In this paper, we use this definition of emotions as it summarizes research showing that emotions are multi-faceted, involving co-existing processes of cognition (e.g., thoughts), behavior (e.g., facial expression changes, activity changes) and physiology (e.g., increased heartbeat, sweating) [42].

At times, people experience different emotions when encountering the same situations. Appraisal theory states that an individual's perception of a situation predicts the emotions that one feels [40]. These perceptions are based on two key factors: *motive consistency* and *evaluation of responsibility* [40]. *Motive consistency* suggests that if an event is consistent with an individual's motives, he will feel a positive emotion; if the event is not consistent, then he will feel a negative emotion. *Evaluation of responsibility* refers to an individual's belief of the cause of the event, specifically whether the event was caused by ourselves, others, or general circumstances. For example, an event that is inconsistent with one's motives and caused by oneself, like if one causes a friend to be hurt, may lead to guilt. On the other hand, an event caused by external circumstances, like if a train is very late, may lead to disgust [40].

2.2 Emotions in student programming

With the recent trends towards studying student affect and motivation, computing education researchers have increasingly studied the emotions that arise in programming [13, 16, 22, 24, 25, 39], recently summarized in [32]. For example, Bosch & D'Mello identified novice programmer affective states, showing that novices experience emotions from engagement and happiness to disgust and frustration [7]. Kinnunen and Simon identified different aspects of freshman programmer experiences, like the "hit by lightning" experience, which occurs when a student encounters a problem that they did not expect [26, 27].

A number of these studies have identified that student emotions correlate with measures of student success, like performance and persistence. For example, Bosch et al. identified that boredom, flow, and confusion are correlated with student performance [6]. Lishinski et al. found that emotions correlate with student performance on projects and has long- and short-term effects on course performance [29]. Haden et al. identified that by using scales of interest, improvement, plan, and satisfaction, they could identify students who are likely to be at risk of failure [25]. Studies have also documented that student emotions correlate with self-efficacy and self-assessed productivity [24, 27, 29]. For example, Kinnunen and Simon documented that after an unsuccessful programming episode, students expressed that they had feelings of inadequacy and stupidity. Additionally, emotions likely influence student persistence through the CS major since enjoyment of programming is a main factor in student decisions to major in CS [28], and students' recollection of past programming assignments are dominated by their emotional experiences [25, 26].

2.3 Existing methodologies for identifying triggers of emotions during programming

A few studies have attempted to identify the triggers of the various emotions that students experience while working on programming problems. Two studies identified a list of triggers for both negative and positive emotions by asking programmers generally about the causes of their emotions [22, 34]. Despite the interview happening directly after a programming session, these studies did not direct the questions about emotions towards specific instances in that session, but were general to their programming experiences. For example, Girardi et al. asked programmers *"What are the causes for your negative emotions during programming?"* These types of questions require students to reconstruct memories based on an association, in this case programming events associated with emotions. Asking students question in this format results in less accurate recollections when compared to asking students to provide detail of specific instances of a programming session [36, 38]. Drosos et al. accessed specific examples of frustration by instructing students to report their emotions on a survey throughout a programming session [20]. While this data provides specific examples of emotional triggers in-action, the programming session is not authentic because the students had to report their emotions throughout the session. Additionally, identifying the points of interest are solely reliant on students to remember to report their emotions. Although they did not directly report on the triggers for student emotions, Bosch & D'Mello addressed these issues by using a qualitative approach to understand the emotions that novices experienced during their first computer programming learning session [7]. The researchers and participants watched the screen recording and front-facing video of the session, pausing the recording at specific interaction events to ask participants about their affect. Since the probing points were predetermined, these findings may be missing instances when students have emotional reactions that the researchers did not expect. In this paper, we explore a new method for studying specific instances of triggers of student emotions in authentic programming episodes.

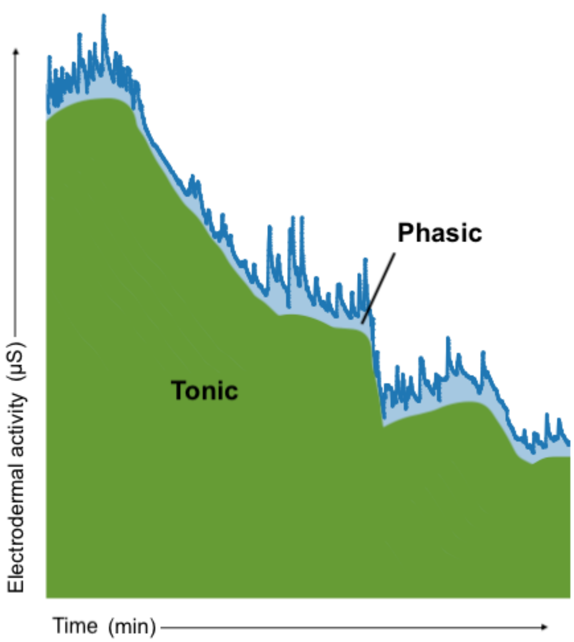


Figure 1: Phasic and tonic activity in an EDA signal.

2.4 Electrodermal activity (EDA) measurement

When a person experiences an emotional stimulus, it creates sympathetic neuronal activity, which results in frequent, tiny changes in sweat production [8, 9, 17, 19]. These changes may not be noticeable to the individual, but can be detected through electrodermal activity (EDA) sensors. Electrodermal activity is the measurement of skin conductance, which is based on the amount of sweat; the more sweat on a person's skin, the higher electrical conductivity.

The EDA signal can be broken into two components, phasic and tonic, seen in Figure 1 [12]. Phasic activity is the short-term fluctuations, or peaks and valleys, that represent neuronal activity. These peaks are referred to as skin conductance responses (SCR) and can be used to study temporally unfolding events, as their intensity reflects physiological significance of events that trigger them [8, 19]. Tonic activity is the general level of EDA and varies slowly, thus is referred to as the skin conductance level (SCL). While SCL is influenced by emotions, external factors can also impact SCL, like time of day [9]. Shifts in tonic level and changes to frequency and amplitudes of peaks in the phasic activity are indicators of changes in emotions [9].

Physiological data, and specifically EDA, has been used to study students' affective state during cognitive tasks [5, 12, 14, 31, 45]. EDA data addresses three common challenges with measurement of student emotions [12]:

- Emotions can occur at any time, but many data collection methods capture information at specific intervals. EDA can be measured continually through an entire activity.
- Participants do not always express the emotions they have experienced. EDA can indicate the presence of emotions, even when participants have difficulty reporting those emotions due to their inability to remember, discomfort in talking about the emotional experience, or challenges with describing emotions accurately.
- Emotions can be subconscious. Not all emotions are conscious to respondents, yet EDA can capture these sub-cognitive reactions [17].

Additionally, EDA data is simple to collect. There are many devices that can measure EDA, most of which have a cost accessible to researchers and only require the user to wear a wrist-band or clip on their finger.

2.5 Studying student programming using electrodermal activity

Computing education researchers have begun to use EDA as a tool to understand student emotions during the programming process. Two studies developed machine learning models that could accurately predict emotions from EDA data of a programming session [22, 34]. To create ground truth data, the researchers periodically interrupted the participants while programming to get self-reported emotions. Then they trained and tested SVM machine learning models using the EDA data. Both studies were able to build models that could reliably predict programmer emotions. While these machine learning algorithms demonstrate the prediction power of EDA data to understand student emotions, machine learning models require large amounts of training and testing data, often for each individual student, as well as development time and knowledge. Additionally,

Figure 2: The Empatica E4 wristband [2].

machine learning algorithms only provide information about the outcomes of the algorithm, and not the context in which the emotions arose, the factors that determined the algorithm's decision or the insights into the transitions between moments.

A number of other studies have also explored the use of EDA in understanding the programming experience [3, 11, 35, 46]. For example, Worble evaluated the number of SCRs that occurred during a programming task and found there was a correlation between number of SCRs and student self-report of their overall emotion [46]. Ahonen et al. used EDA to look at emotional synchrony between pair programmers [3]. They utilized visual inspection and signal evaluation to investigate differences and similarities in pair programming roles.

In this paper, we present to the ICER community a new approach to utilizing electrodermal activity (EDA) to provide insights into the events that trigger students to experience emotions while programming. Additionally, we provide a new, accessible method that computing education researchers can use to understand student emotions by harnessing the power of EDA through qualitative analysis.

3 METHOD

This study aimed to answer our two research questions by identifying the events that trigger student emotions while programming and studying whether students' remembered experiences align with EDA data. We designed a lab study in which we collected EDA data while students worked on programming problems, and then conducted a retrospective interview that leveraged the EDA data using a cued-recall technique to improve students' ability to remember their emotional experiences. We chose to analyze our data using a qualitative methodology to uncover the rich context of students' emotional experiences while programming.

3.1 Participants & setting

We recruited 14 undergraduate students (10 men and 4 women, aged 18-21) from a mid-sized private university in the Southeastern United States. We conducted this study at the beginning of a semester, in January 2022. All participants had only completed one introductory CS course; some participants had started a second introductory CS course. We choose this population of students because we believed they had enough experience with CS to develop perceptions and opinions, but were still deciding if they should pursue CS. We recruited students through emails sent by the professors of the introductory programming courses and announcements made in class.

3.2 Study procedure

The first author conducted two-hour interviews with participants individually, following social distancing and masking protocols. There were two sections of the interview: a programming session and a retrospective interview.

3.2.1 Programming session. Participants were directed to work on a programming problem for 30 minutes. The problem description, which described the expected functionality and provided examples, asked participants to write a function that removed duplicate words from a sentence, where duplicates were case-insensitive. We designed the lab study to emulate a normal programming session as much as possible. The researcher instructed participants to use resources as they would for a normal homework assignment. Participants used their own laptops, but used the jGrasp IDE [18], which was new for many participants. When the participant began working on the problem, the researcher left the room. While programming, participants wore an Empatica E4 wristband EDA sensor [2, 33] on each wrist, as seen in Figure 2. We choose to use a wristband sensor instead of fingertip sensors because it is less disturbed by the movements involved in using a keyboard and mouse.

3.2.2 Retrospective interview. After the programming session, the first author followed the procedure in Section 3.3 to analyze the EDA data captured by the Empatica E4 device. This analysis produced a list of timestamps of skin conductance responses (SCRs) captured in the EDA data, which indicates potential emotional responses.

The first author then conducted a retrospective interview [21] using cued-recall techniques [4, 10] with the SCRs as cues. The researcher and participant watched the screen recording and laptop camera recording of the programming session together, with the list of SCR timestamps displayed onscreen as a reference. As they watched the recording, the researcher asked the participant to describe their programming experience and any emotions they felt. When they reached a point in the recording that aligned with an SRC timestamp, the researcher informed the participant that an SRC had occurred and asked them to describe what happened at that moment and whether they experienced an emotion. Any time the participant identified an emotional reaction, the researcher asked whether it was a positive or negative emotion, and what they believe caused the emotion.

One potential concern with this approach is that participants might feel pressure to create a narrative to fit the SRCs, for example due to a bias caused by the study's demand characteristics [37]. To address this issue, the researcher discussed the nature of EDA data with participants at the beginning of the interview, before watching the recording of the programming session. Specifically, the researcher told participants that while EDA sensors are good at detecting when emotional reactions occur, other unrelated factors can also cause SCRs, such as rapid arm movements. Our goal was to ensure that participants felt comfortable sharing that they did not experience an emotion, or experienced an emotion due to a distraction rather than the programming task, at a moment when the EDA sensor picked up an SCR.

3.3 SCR detection

The first author conducted the analysis for determining the SCRs in the programming session after the participant finished programming but before the retrospective interview. To identify SCRs, the researcher first used the peak-detection algorithm from EDA explorer [43]. To determine the correct parameters to set the tolerances for the algorithm, the researcher conducted a parameter sweep with 16 sets of parameters. The parameters included a tolerance value, or the minimum amplitude that a peak must reach in order to be considered an SCR. The researcher chose the parameter set that best identified distinctive peaks without capturing noise based on visual inspection of graphs of the detected peaks from each parameter set.

Finally, the researcher visually inspected the graph to remove any mislabeled SCRs and add any unlabeled SCRs in order to improve the accuracy of the results. Visual inspection has been used to identify SCRs in prior work [12, 35]. Visual inspection can improve the accuracy of EDA data because SCRs may not always be perfect peaks. For example, when there are multiple SCRs in a row, the second peak might start before the first resolved, making it difficult for an algorithm to identify them. The result of this analysis for participant 12 is displayed in Figure 3.

3.4 Identification of triggers of emotions

We conducted a qualitative analysis of the interview data to identify types of moments in the programming session that triggered emotions. The first author generated an initial codebook for moments that trigger emotions. She started by reviewing a subset of the interview transcripts with an open-coding protocol [15], focusing on instances when participants expressed an emotion and described a trigger for that emotion. She iterated on the themes as she reviewed more transcripts.

This process generated a list of emotion triggers and associated descriptions of the triggers broken into positive and negative emotion categories. For example, after having an issue with the compiler a participant said: *"I think I was happy that ran"*. We categorized the trigger of the positive emotion in that moment as "resolving interface issues".

The first two authors then used the codebook to independently code emotion-trigger pairs in the data. They coded interviews separately and discussed discrepancies after coding each transcript, iterating on the codebook when necessary. To check consistency, the two authors independently coded three transcripts, or 21% of the data. They had a percent agreement of 83%, which represents good agreement. The first author coded the remainder of the transcripts.

3.5 Analysis of EDA data with respect to student experiences

We next investigated how the EDA data reflects student experiences throughout the programming session to better understand how EDA data can be used to interpret the student programming process. We aligned the graphical representation of the EDA data with participants' descriptions of their experiences across the entire programming session. While in the interview we only used the timestamps of SCRs, for this analysis we considered three features to reflect more gradual emotional changes that take place across time. The three features that we used to analyze the EDA graphs are: the amplitude of SCRs, frequency of SCRs, and drifts or changes in skin conductance level (SCL), as they are most indicative of emotions [9].

During the analysis, the first author annotated each participant's EDA graph with the student's recollections of the activities they performed and the emotions they experienced during the programming session. This side-by-side view of student report and physiological data facilitated analysis of how EDA evolved with students' emotional reactions as they worked through the problem. Then the first author wrote memos about the patterns that she observed in the EDA data, observing how the descriptions of the events aligned with changes in the EDA features [41]. The memos were used to detail the first author's observations of the notable behaviors of the EDA data as related to the student recalled experiences. Finally, the first author reviewed the memos and the annotations to identify common patterns seen across multiple participants.

4 FINDINGS

We share our findings, corresponding to our two research questions, in the following two sections. First, we present a set of events that trigger students to experience positive and negative emotions while programming and present detailed examples from our data set. Second, we present evidence of the relationship between student remembered experiences and patterns in their EDA data, identifying a few common emotional patterns that arose across multiple participants.

4.1 Events that trigger student emotions

To answer our first research question, we identified a set of programming events that students said triggered an emotional reaction. From the qualitative analysis, we identified 21 themes of these events, including 8 that participants reported caused positive emotions (see Table 1) and 13 that participants reported caused negative emotions (see Table 2). When identifying the triggers, we only labeled moments where both the emotion and trigger were present. Thus, there may be other instances of these triggers where participants did not experience an emotion. The count represents the number of instance when these triggers caused an emotion. Interestingly, many of the positive events were opposites or counterparts to

Table 1: Events that triggered positive emotions during the programming session.

Trigger	Explanation	Count
Getting direction from a resource	When a student learns something or finds something useful in a resource.	22
Typing code / making progress	When a student has positive emotions from the action of typing code in the editor. This occurs when there is the feeling of being productive.	11
Completing a step in coding problem	When a student completes a task, whether it is a step in the programming problem, or a subgoal in their process to completing the problem. This is outcome oriented and denotes concrete completions of a step, generally evidenced by running code, but not always.	10
Having a plan	When a student has a plan for their code. This also could be relief or excitement from recognizing a new plan.	9
Fixing errors	When existing errors are fixed and no longer present.	7
Remembering syntax correctly	When a student remembers syntax without external help. This is not about writing code, but specifically remembering without help.	5
Understanding the problem statement	When a student has positive reaction from reading the problem statement because they understand the question.	3
Resolving interface issues	When a student resolves an issue unrelated to solving the programming problem. Some examples include: IDE, finder, resource access. This is the end of the “interface issues” trigger.	3
Other	When a student experiences a positive emotion but the triggering event does not fit into one of the specified labels.	4

Table 2: Events that triggered negative emotions during the programming session.

Trigger	Explanation	Count
Not knowing something	When a student feels they need to use a resource because they do not know something, because they forgot it or do not know it.	23
Interface issues	When a student has issues unrelated to solving the programming problem. Some examples include: ads, IDE, finder, resource access.	15
Resource not helping	When a student uses a resource but does not find it to be helpful. This may occur because they do not understand the resource or can not find the right source.	15
Realizing there is an error	When a student gets an error. This is about the existence of an error and not the struggle with fixing the error	14
Struggling while trying to fix an error	When a student struggles with or spends a long time working on errors, whether they are simple or not. For example, repeatedly trying to solve the same error multiple times unsuccessfully.	12
Intimidation from reading problem statement	When a student has a negative reaction to reading the problem, either because they do not know how to approach the problem or because they are intimidated by it.	11
Not making progress	When a student feels they are not making progress towards the solution. This could be displeasure about spending time thinking or not making progress on the problem.	6
Realizing code/plan not working as expected	When a student has implemented code and realizes the code or their plan for the code is not working in the manner they expected. This occur when running or from reading the code.	4
Changing approach / deleting code	When a student changes the approach they have been taking towards the problem. This often appears when a student deletes lines of written code.	4
Not understanding error message	When a student does not understand the text of an error message.	2
Struggling to fix program behavior	When a student struggles to fix a logic error or incorrect code behavior.	2
Encountering code formatting issues	When a student has issues with the formatting of the code, not the functionality of the code.	2
Not remembering problem description	When a student has to read the problem statement again because they forgot it or misunderstood it.	2
Other	When a student experiences a negative emotion but the triggering event does not fit into one of the specified labels.	5

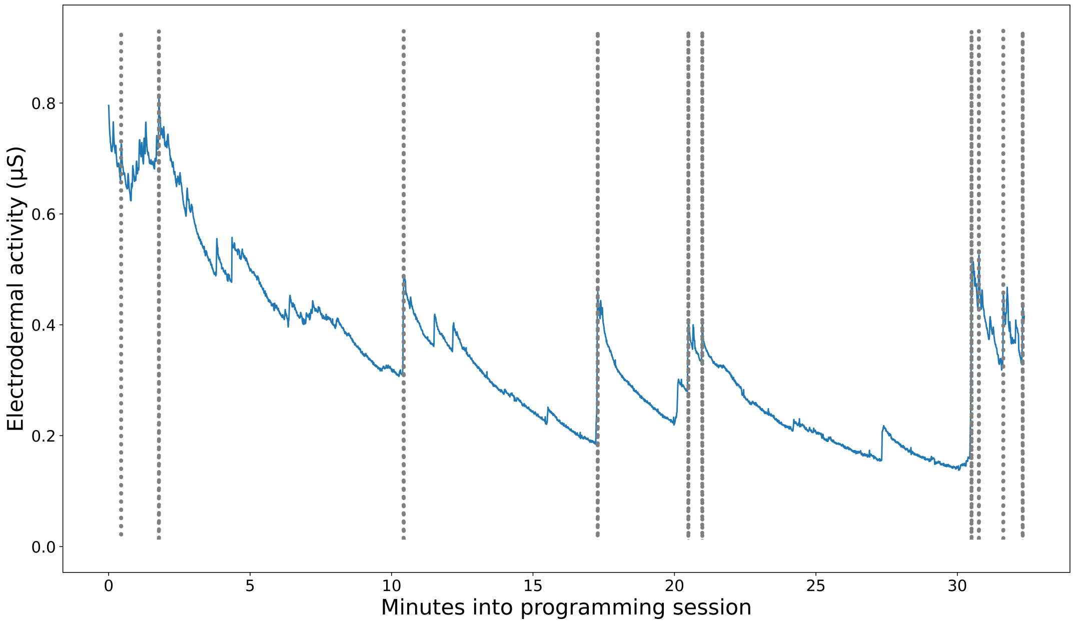


Figure 3: Results from skin conductance response (SCR) detection for P12. SCRs are marked by dotted vertical lines. EDA level is displayed in blue.

negative events. For example, we found that negative emotions occurred when a student experienced issues working with their tools (code: "interface issues"), and positive emotions occurred when a student relieved those issues (code: "resolving interface issues").

Many of these moments involved interaction with a variety of information, resources and tools, like reading the problem statement, interacting with the console in the IDE, and searching for resources online. While students may interact with the same tool or process, we often found that they experienced different emotions depending on their intentions, responses, and results of the interaction. For example, students often searched for resources to help them complete the coding problem, but their emotional responses differed depending on whether they found the resource helpful (code: "useful resource"), unhelpful (code: "resource not helping"), or difficult to use (code: "interface issues"). We also found that students had different reactions to the exact same circumstances. For example, some students had a negative emotional reaction when they could not remember syntax and used a resource to look it up (code: "not knowing something"). However, other students in the same situation did not experience an emotional reaction at all.

While we did not reference the physiological data while conducting this analysis, that data informed the retrospective interview. One risk of this approach is that participants may have felt a need to explain the SRCs, even if they did not experience emotions at those moments, which could lead to inaccurate trigger-emotion pairs. To confirm that participants felt comfortable sharing that they did not experience an emotion, or experienced an emotion due to a distraction rather than the programming task, we reviewed all instances where the researcher asked the participant if they experienced an emotion when an SRC was detected. We found that all 14 participants said that they did not experience an emotion at least one time when directly asked. For example, when asked if he experienced any emotions at a particular SRC, P3 said *"I think not yet. I'm not sure what it would be yet."* Many students also shared that external factors impacted their programming session. For example, when asked if he experienced an emotion at a particular point in the programming session, P10 said *"No, I don't think so. I just started texting on my phone."* These responses provide some confidence that students were accurately reporting their emotional experiences in the interviews.

The benefit of this methodology is that it allowed us to question students about very specific moments during the programming session. As a result, we were able to dig into the specific experiences that led to emotions. In the following sections, we describe nuances and overarching themes for some of our most frequent codes, highlighting differences in the ways that students described the triggers of their emotions.

4.1.1 Not knowing something. "Not knowing something" was the most frequent trigger for negative emotions in our data set, with 23 instances across 11 of our 14 participants (79%). The emotions that students most frequently described at these moments were frustration (7 mentions) and annoyance or irritation (8 mentions). Some students also associated this moment with embarrassment or shame (4 mentions).

When describing the moments of "not knowing something", participants frequently mentioned that they had once learned the content they were trying to remember. *"I know I've written this type of code a lot of times. So I'd say I was probably annoyed with myself for not remembering it...that it didn't come in my head straight away"* said P3, when describing his annoyance for not remembering the syntax for using the length method. P10 described his lack of memory for how to separate words in a String, saying *"It was annoying because I had done that before, I just couldn't remember"*. The memory that this information was once at their fingertips seemed to contribute to their irritation that it was now gone.

Participants also described the simplicity of what they had forgotten as a reason for their negative emotion around forgetting. *"I was a little bit ashamed that I forgot how to do something so simple in Java,"* said P11, describing her memory lapse when importing the Java Scanner. P5 expressed that he "should" recall the typical first line in a Java program: *"[I]t was muscle memory back when I took that class, where I would just import* `java.util`*, and then I forgot how to spell it."*. The participants perceived these minor details as basic, and thus within their capacity to memorize.

At times, participants generalized their lack of knowing something quite broadly, stating that they had forgotten everything about a topic or even a whole course. P1 said, *"most of my knowledge from the class I might've forgotten, so going back and having to use a resource was a little bit irritating, certainly."* P13 recalled feeling "embarrassed" after watching the screen recording of getting errors in her code. She reflected, *"I realize[d] I don't remember any of this"*. Upon looking at her previous assignments, P8 described her emotional response: *"okay, this is harder than I thought. I don't really remember anything from past semester. So I would say again, some panic"*. It is unlikely that these students forgot all of the course content, but in their emotional state they felt a substantial lack of knowledge.

In each of these instances of the "not knowing something" code, we saw that self-judgment was present along with the negative emotional response. When these participants did not know something or had to look something up, a negative assessment of their own ability was often a key factor in their emotional appraisal of the moment. This aligns with past studies that found correlations between self-efficacy and negative emotions [27, 29]. This also aligns with past work by Gorson & O'Rourke that found that students negatively self-assessed when they needed to use a resource to look up syntax or research an approach [23].

We also found that students often had the opposite reaction when they did know something. Specifically, we found five instances where students experienced a positive emotion after "remembering syntax correctly". For example, after P3 wrote the initial structure of his code, he said *"I was pretty happy that I did remember it on the first try, which was cool."* This indicates that student emotions are often tied to whether or not they can recall coding content while programming.

4.1.2 *Getting direction from a resource.*

"Getting direction from a resource" was the most frequent trigger for positive emotions in our data, with 22 occurrences across 11 of our 14 participants (79%). When participants mentioned feeling a specific emotion after finding a resource helpful, they were most likely to describe relief or relaxation (6 mentions), happiness or joy (4 mentions), and excitement or hope (3 mentions).

We found that sometimes, simply the realization that they had access to a helpful resource was enough to trigger a positive emotional reaction. P8 found relief when she realized she could search the internet for help despite not yet identifying specific content to use in the program at hand. She said: *"It's like, oh, okay. Yes. I can use Google. I think it's a sense of relief, like okay, I have more sources. I can use other people's ideas. I just remembered that."* Similarly, P2 recalled that his ZyBooks content might be a useful place to find insight on the problem, leading to happiness. He described the triggering moment as, *"thinking about going onto my class content, which I guess to some extent it relieved me from me being lost."* In both cases, the possibility of finding something helpful in a resource was enough to spark positive emotions.

Participants often had a specific goal in mind when looking through resources, and thus experienced positive emotions when they found it. P4 described such a scenario as she looked through past programs on her computer to find out how to use the `.indexOf()` method: *"I knew what I wanted, but I didn't know the vocabulary for that and I just found it on my notes."* P14 reported a similar situation, saying she felt *"a little bit of excitement that I saw a result that was what I was looking for"* as she searched Google for information about how to split Strings.

Other times, participants looked through resources in a less targeted manner, and thus when they chanced upon information they deemed helpful they experienced a positive emotion. After P12 typed a broad Google search, he recalled a moment of joy. He said: *"this* `.split()` *method did not come to me before, in my mind. And now I just saw it randomly. So I thought maybe this would work."* P8 had a similar experience when using Google. She recalled positive emotions when looking through a Stack Overflow page: *"I saw* `.length` *and my thought process was of course you can do it with that* `.length`*."* These participants had a chance encounter with a useful clue, resulting in a positive feeling.

We also saw the opposite reaction to occur when students were not able to find the help they needed from a resource. We found 16 instances of negative emotions when participants encountered moments of "resource not helping". For example, when P1 was not able to use a website to solve his confusion, he stated *"this one irritated me so much, just opening this website. I was not having it. It was no help ... I was annoyed."* P13 had a similar experience, stating that he was frustrated when he was *"not finding what I was looking for."* These findings suggest that resources can often be a source of emotional stimulus depending on if students are able to find the information they need.

4.1.3 *Typing code / making progress.*

"Typing code / making progress" appeared as a trigger for positive emotions in 11 instances across 6 of our 14 participants. Participants expressed happiness or general positivity in these moments (6 instances), as well as a sense of accomplishment or confidence (2 instances) and relief (2 instances). Participants described these moments in terms like "making progress" (P2, P14), "moment of progression" (P12), and "getting/going somewhere" (P1, P8, P14).

When asked about why they had a positive emotion in these particular moments, participants often described the process of writing code. For example, when P2 was recalling a moment when he erased erroneous code partway through the coding session, he said, *"I pushed delete and I'm thinking of actually progressing"*. He described his emotion in that moment as *"happy that I'm starting to write something"*. For P14, her feeling of hopefulness was triggered by *"seeing the program start to come together and adding more things to it"*. P12 described the precise moment when he felt positively about his progress as when he typed code in his editor: *"When I start writing* `int i` *and* `int x` *is equals to one, for this length, I feel a moment of progression, that I'm actually going forward. I'm actually going forward in this problem towards the solution"*. These participants associated the action of adding content to the editor or the anticipation of typing code with a positive sense of progression, even though they had not yet run their code to test whether it truly worked.

The opposite was also true: negative emotions were associated with "not making progress", which were often tied to not writing code or typing. "Not making progress" was a trigger for negative emotion for 4 participants, across 6 instances. Participants used a variety of terms to describe their feelings in this scenario, including "anxious", "worry", "frustration", "stress", and "distress".

P1 expressed that spending too much time thinking and not starting to code soon after reading the problem statement led to a feeling of "anxiety". He said: *"I was processing it and I realized I was taking a little bit too long to start coding. So I was like, 'Shoot. Okay, let's go. Let's transfer over to start coding'"*. Even in the beginning of the programming session, P1 felt negatively if he was not implementing code.

At times, the emotions around making progress were so strong, they determined the participant's behavior. For example, P2 was using the internet to get help on the problem, but did not find the answer he was looking for. He felt frustrated *"because I'm not actually progressing in what I want to do"*. Instead of continuing to use resources, he went back to his code and started typing. He said,

> *I know one thing that I did wrong, which is just go back to the code. I should have known that I'm not going to just figure out suddenly something off of just going right back into the code and doing something. I just really felt like I wanted to progress...I wanted to just I guess, still maybe delete something or maybe type something, but it's not really even useful. So, I just wanted in any way, having the feeling of progressing.*

His desire to have the positive feeling of making progress was so strong that he stopped and went back to the code, even though he admits that it was futile.

We found that our participants' sense of progress can drive emotion, both positively and negatively. This sense of progression is often associated with the action of typing code in the editor, whether it is beneficial or not and can drive participant programming behaviors.

4.1.4 Realizing there is an error. We found that the moment when participants realize that an error is present can yield negative emotions. We identified 14 of these instances across 7 of our 14 participants (50%). These participants expressed specific emotions like frustration or annoyance (6 mentions), worry (2 mentions), and disappointment (2 mentions) upon realizing there was an error in their code.

Participants shared that one reason "realizing there is an error" can trigger emotions is that it revealed that their code was not correct. "Realizing there is an error" was more likely to cause negative emotions when participants expected their code to be successful, versus when they did not expect it to work. P1 described such a moment: *"The main thing about the error that frustrated me was that I thought I had a solution... but I just didn't"*. Similarly, P7 described his disappointment at seeing an error message after compiling his code as: *"It's just a let down, as I said. It's a confidence roller coaster. Hitting that compile button, I knew that it wasn't going to be completely correct, but yet I had the confidence it was"*. Although he was aware that it was unlikely that his solution would work, he still believed it would, and thus was saddened by the error.

Even when the error was relatively minor and easy to fix, some participants still felt negative emotions. For P6, a minor error caused a negative emotion because it was a repeated error. He described his reaction as: *"I got the same error I got before. So I knew how to fix it, but then I had another error, but it was something really simple. So probably just another slight frustration"*. While the frustration was slight, P1 had a negative emotion after experiencing multiple errors in a row. P5 had a similar experience, where he felt annoyed at a small error after receiving multiple, even though the error itself was understandable and easy to fix. He said, *"I guess that would've been an instance of having screwed that up a couple of times and then again, making another tiny little mistake"*.

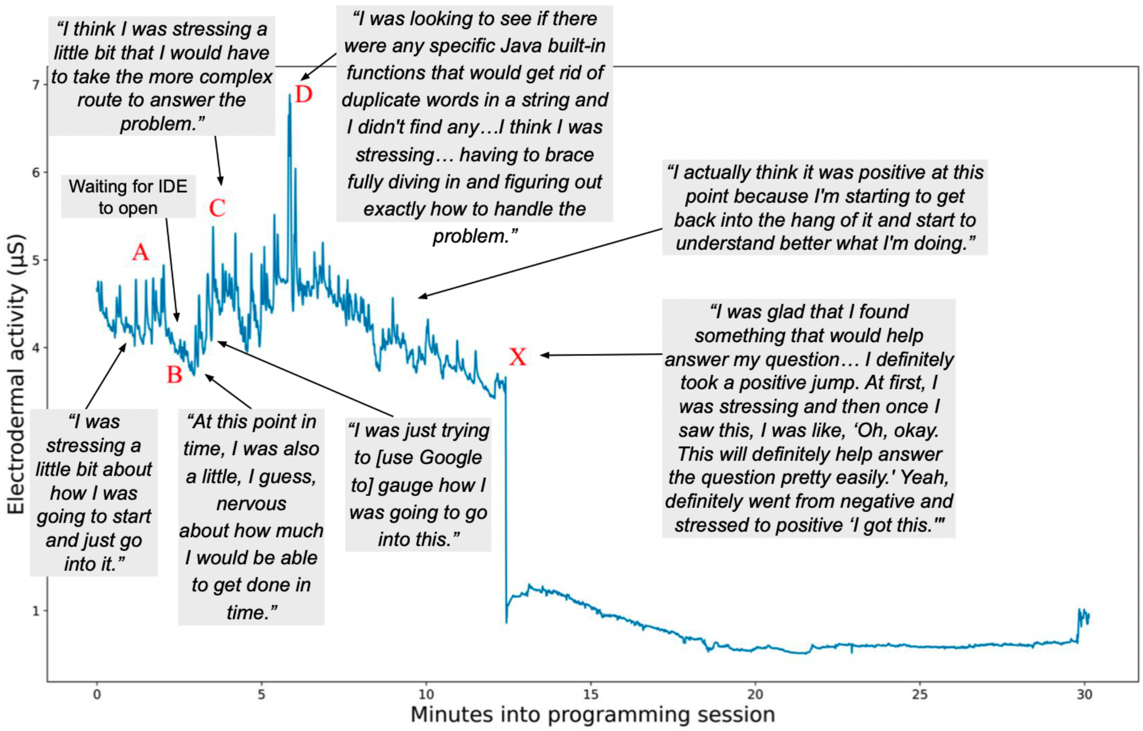
Just the recognition that an error exists caused a negative reaction for some of our participants. At times, these errors indicated large issues in their plan or code, while on other occasions, the errors were quite simple. This may be due to the circumstances in the programming session, like finding multiple errors in a row, or negative views on errors.

4.2 Emotional experiences reflected in EDA data

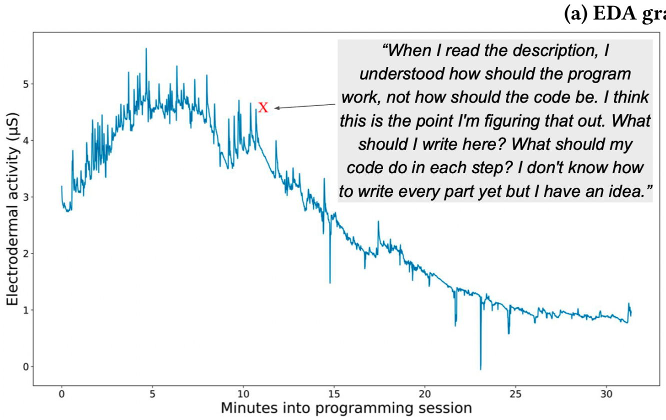
To answer our second research question, we explored the relationship between the participants' remembered experiences and the tonic and phasic changes in the EDA data, finding strong alignment. Specifically, the EDA data provides a map of student experiences, not only highlighting emotional moments (reflected by peaks), but also aligning with their emotional state over time (reflected in tonic level and peak frequency and amplitude). We share two participants' programming episodes to demonstrate how EDA data reflects their experiences. We also share three patterns that we observed across participants in the analysis process.

4.2.1 Participant 11. We start by describing the programming session of Participant 11 (P11) and the aligning EDA data. See Figure 4a for the respective EDA graph with associated quotes. At the beginning, P11 described feeling nervous and stressed about working on the problem. This initial reaction to the programming problem aligns with the small but noticeable peaks that we see around **Marker A**. Following **Marker A**, we observe a downward slope in the EDA data at **Marker B**. At this point, P11 is waiting for a resource to open. From the interview data alone, we would not know if P11 continued to feel stressed during the waiting time or was just sedentary. From the reduction in peaks and downward slope of skin conductance level (SCL) in the EDA data, we can infer that the participant was sedentary in this transition. Once P11 has access to the resource, she reported feeling nervous and anxious about how much she will complete. This nervousness occurred around **Marker C**, when the SCL rose and the peaks increased.

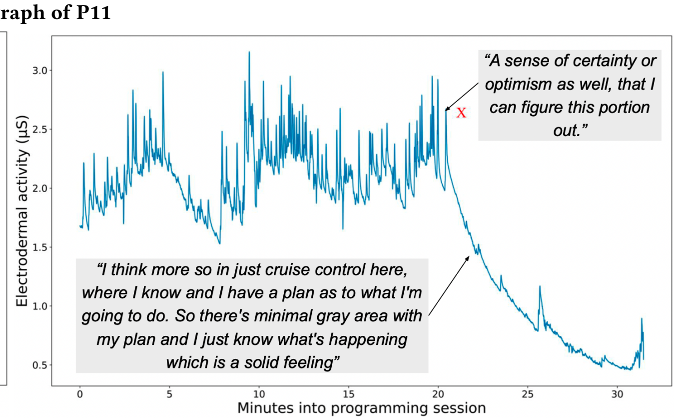
After searching through resources, P11 realized that the preset function she was looking for did not exist, and she would have to write the code for this sub-task herself. This caused a strong



(a) EDA graph of P11



(b) EDA graph of P8



(c) EDA graph of P9

Figure 4: EDA graphs of participants demonstrating the "Cruise Control" pattern, which begins on each graph at the Marker X.

emotional reaction, which lines up with the large spike that can be seen at **Marker D**. Afterwards, she continued to look for a resource to help solve the problem. The SCL drifted down as she made progress in understanding the problem and determined an approach. Around the time of **Marker X**, she found a website that helped her develop a concrete plan for implementation. After that point, she was much calmer. Her energy shifted from determining how to solve the problem to implementing the plan laid out in the resource.

4.2.2 "Cruise Control". For a number of participants, we found that a large decrease in EDA (both SCL and phasic activity) occurred at the same time as the participant determined their plan and started to implement their code. This occurred for P11 at **Marker X**, as described in the previous section. We also saw this phenomena with P8 and P9, shown in Figures 4b and 4c. These participants described an event that helped them determine their plan for approaching the problem, allowing them to shift their focus from research and planning to implementing the new approach. The participants described feeling calmer and less stressed while implementing compared to planning, because they perceive implementing as a more confined problem space then planning. P8 described why she was less stressed after making a plan: *"the first part was more frustrating and the second part was more exciting than stressful because as* *time passed, I had a better understanding of the problem and better understanding of how to use the resources and got comfortable with the situation more".*

The EDA data for this point of the programming session demonstrated very strong similarities across the three participants. At each of these moments, there was a steady, large drop in EDA followed by consistently lower SCL and phasic activity. This can be seen in Figure 4, where we have marked an X at this moment in each of the graphs for P8, P9 and P11. P9 used the term "Cruise Control" to describe the implementation phase. He said: *"Yeah, I think more so in just cruise control here, where I know and I have a plan as to what I'm going to do. So there's minimal gray area with my plan and I just know what's happening, which is a solid feeling."* His description of "Cruise Control" aligns with the other students' experiences, as they all expressed that they felt that the problem was mostly solved after the turning point, and was evidenced in the EDA data.

4.2.3 Participant 1. Next, we describe P1's programming session and respective EDA data. See Figure 5a for the respective EDA graph with associated quotes. Shortly after beginning, P1 reported being disappointed when he realized that he was spending a long time thinking and had not begun to code. He expressed that he was anxious and thought *"Shoot. Ok, let's go. Let's transfer over to start coding"*. This feeling came with a *"sense of urgency"* to start implementing, which aligned with a peak in EDA at **Marker A**. Later in the programming session, P1 tested his code and got an error. The error aligns with the peak in EDA at **Marker B**. After the error, the annoyance continued as the participant could not determine how to fix the code despite using resources. This aligns with the section following **Marker B** where there continues to be frequent peaks in the EDA data. He then recognized that he was unsuccessful at using resources around **Marker C**, expressing disappointment that aligns with a time when the EDA peaks have high amplitude and are close together. Both the density and height of the peaks and the participant's qualitative description indicate that this shaded section was a strongly emotional and frustrating part of the session.

P1's SCL drifts downward as his emotions became more positive. He shares that he actively calmed himself down. He describes that he was *"attempting to gather my thoughts and bring my emotions down to a level where I could actually make good and viable steps towards working on a solution."* Then, he recalled feeling calmer and starting to make progress around **Marker D**, where there is a decrease in SCL and less frequent peaks. The negative emotions reemerged at the end of the session. He described his thinking as: *"being reflective, but in a bad way. More as in, 'Wow. Well, I suck'"*. This negative emotion and evaluation aligned with **Marker E** on the EDA graph, which shows a large, rapid increase in EDA followed by a few additional peaks.

4.2.4 High emotion sections of a programming session. A visual inspection of P1's EDA graph shows an increase in physiological activity around Markers B and C with all three indicators of physiological reactions: frequent peaks, increased altitude of peaks, and increased SCL. As expected, this aligns with his description of particularly strong emotions, indicating that he reached a peak of frustration. Looking across the EDA graphs, we found other participants with a section of the programming session that has multiple indicators of physiological reactions and aligns with a particularly emotional section of the programming session, like P3 and P10, which can be seen in Figure 5. These sections may be particularly important to identify and understand because participants are having such a significant physiological reaction to the programming experience in these moments.

P10 had a section of high frustration and annoyance because he did not have the resources he wanted and was not able to remember how to do part of the programming problem. This frustration directly aligned with peaks in the EDA that had significantly higher amplitudes than the earlier section of the problem and slight increase in SCL, seen at the beginning of the shaded section of the graph. The peaks persisted as P10 could not find what he was looking for. Relief from this highly emotional part came when he found a resource. The peaks subside in the EDA data and the SCL began to lower at this time, right after the shaded part of the graph. In the identified part of the programming session, P10 experienced strong emotions, aligning with high amplitude and frequent peaks in the EDA data.

In the EDA graph of P3, there is an increase in frequency and amplitude of the peaks (at the first part of the shaded section) followed by an increase in SCL (at the second part of the shaded section), which aligned with a particularly frustrating part of the programming session as P3 struggled with an error. Following that initial error, the participant continued to struggle with errors. Around the end of the shaded part, the participant was feeling even more frustrated. He described this as: *"I'd say that coming up here was probably the most frustrating part of the whole thing, like when this doesn't work."* At the time when there was an increase in SCL and high peak amplitude and frequency, P3 describes the highest frustration.

4.2.5 Low EDA activity indicating fewer negative emotions. While the majority of the participants had both tonic and phasic activity in their EDA data throughout the programming session, two participants, P4 and P14, had steady EDA levels throughout the entire programming session. Specifically, the graphs for P4 and P14, seen in Figure 6, show consistent EDA during the programming problem, both in SCL and in the lack of phasic activity. For both participants, this sharply contrasts their EDA activity before the programming session, in the shaded region, which starts with high SCL and has phasic activity, demonstrating the range their EDA could reach if they encountered emotional stimuli.

During the qualitative coding analysis, we identified the fewest instances of negative emotions for these two participants. P4 and P14 had 2 and 4 negative emotions respectively, while the remainder of the participants had an average of 9.25 negative emotions, ranging from 6-15. P4 explained why she did not get an emotion when she realized there were errors, a trigger for some of the other participants. She said: *"Every time after I type code and I run it for the first time, I expect it to fail. So that's why ... it didn't affect me that much either way."* Similarly, when asked about emotions, P14 expressed that he was mostly thinking and not experiencing emotions. For example, after running his code, he said *"I think it's still mostly just kind of thinking and sort of analyzing, but the program*

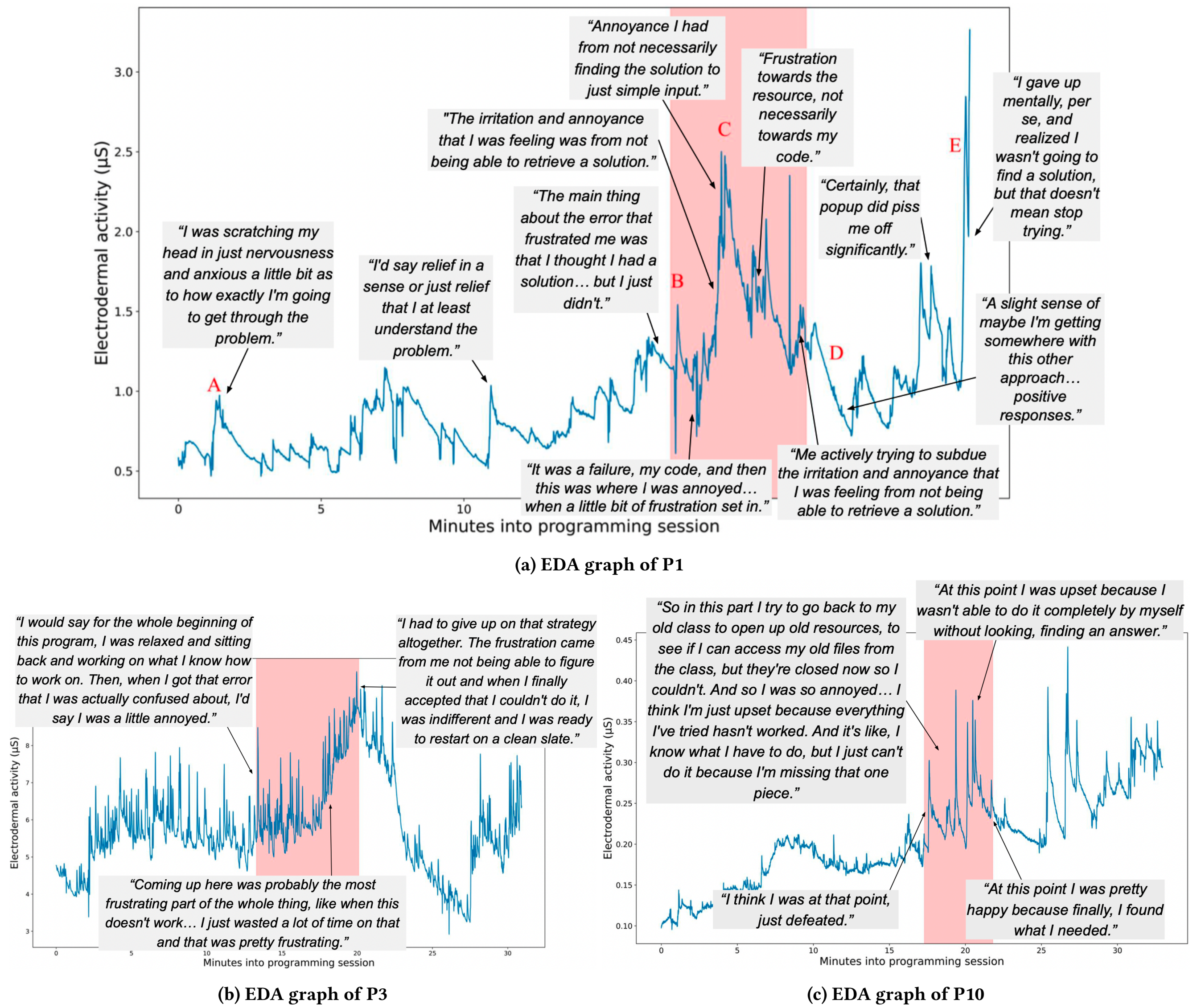
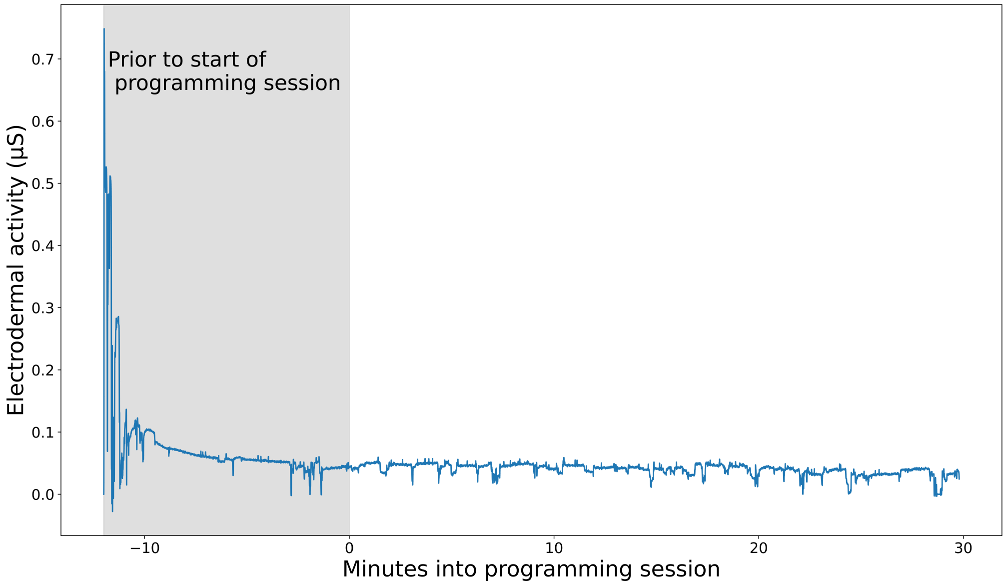


Figure 5: EDA graphs demonstrating high emotion sections, indicated by the shading.

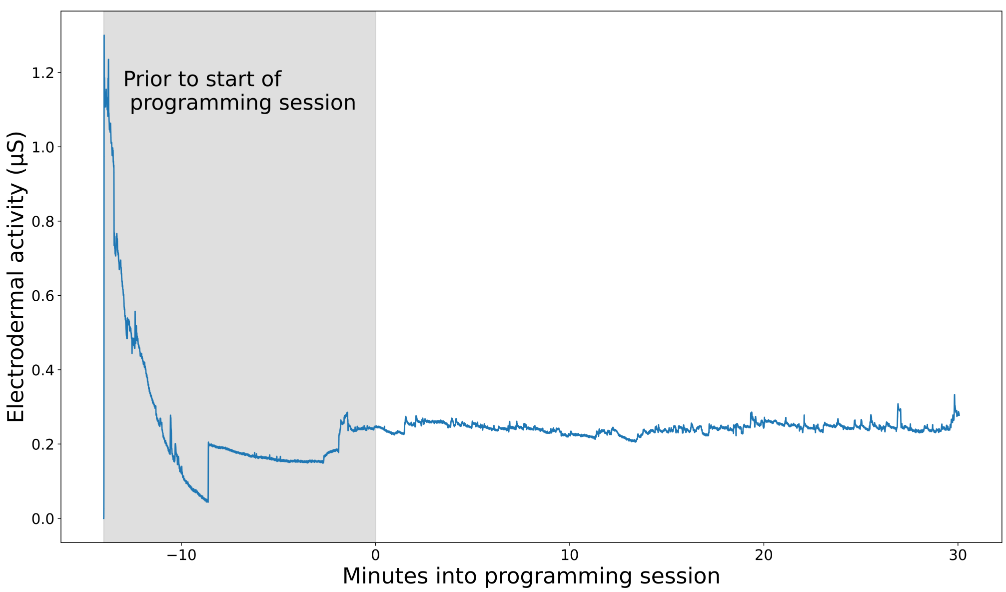
just outputs. Trying to think about what to do next." Interestingly, both P4 and P14 had similar number of positive emotions to the rest of the participants, 6 and 9 respectively, while the remainder of the participants had an average of 5 positive emotions, ranging from 0-9. In our data, positive emotions seem to not be as influential to phasic or tonic activity in EDA as negative emotions.

5 DISCUSSION

Our analysis of retrospective interviews informed by EDA data from 14 undergraduate students allowed us to identify 21 different events that trigger emotional experiences while programming. These triggers and associated emotions provide new insights into the programming experiences that are most salient for students. The most common event to trigger a negative emotion was the experience of not knowing something, highlighting the pressure that students put on themselves to remember syntax and problem-solving approaches they have used previously. Similarly, the most common event to trigger a positive emotion was getting direction from a resources, showing that finding direction and having a plan is both relieving and exciting. Interestingly, many of the triggers of positive and negative emotions were opposites of each other, which can help us identify the big moments that are likely to produce emotional reactions, such as making or not making progress and getting or fixing an error. While all but one participants experienced positive emotions, all participants experienced negative emotions and we saw a larger number of negative emotions overall. Furthermore, many participants exhibited strong reactions to



(a) EDA graph of P4



(b) EDA graph of P14

Figure 6: EDA graphs demonstrating participants who experienced few negative emotions while working on the programming problem.

negative triggers, for example by jumping to self-judgment, which may suggest underlying issues in programming confidence.

These insights into students' emotional experiences while programming have important implications for research and practice. Previous research shows that experiencing negative emotions while programming correlates with lower project and course performance, both in the short and long term [30]. As a result, the finding that many students experience more negative emotions than positive ones while programming is problematic, particularly since some students show indications of negatively self-assessing in response to emotionally-triggering events. Future research could explore the relationship between emotional reactions and other key factors, such as self-efficacy [27, 29], sense of belonging [44], and negative self-assessment moments [23] to better understand the implications of negative emotions while programming. In the near-term, our findings on the common triggers of both positive and negative emotions could inform pedagogy and practice, for example by prompting discussions about common emotional experiences and discussing strategies for managing feelings of frustration, anxiety, or under-confidence while programming.

Our second analysis identified broader patterns in the EDA data that aligned with participants' remembered experiences. We found that the EDA data clearly reflected the narratives that students described in their retrospective interviews, even though participants did not have access to the full EDA graphs. Common patterns, such as a reduction in tonic SCL in the EDA data after devising a plan and beginning to implement it, arose from the data despite our small sample. Furthermore, we found that students can have very different emotional experiences while programming. Some students experience very few emotions, reflected in their steady and low EDA data, while others experience periods of strong and intense (usually negative) emotions, which aligned with high phasic activity and a raise in SCL. These initial findings suggest many directions for future research. For example, researchers could use patterns in EDA data to compare student experiences or isolate particularly interesting segments of a programming session. Further studies could also use this methodology with a larger sample of students to identify additional common patterns in student programming experiences, and potentially study the relationship between these patterns and outcomes like performance and self-efficacy.

This research was enabled by a new methodology that combines EDA data with retrospective interviews. Our findings suggest that EDA data can serve as a valuable resource for prompting student recollections during interviews. Students easily identified and described the emotions they experienced at SCRs, along with the associated context in their programming session. They also felt comfortable telling the researcher when no emotion had occurred at an EDA peak, suggesting that their recollections of emotions were valid. As a result, we believe this is a promising method for gaining insights into students' emotions while programming. In the future, a similar approach could be used to explore emotional reactions during many types of coding activities such as reading unfamiliar code, using instructional materials, or pair programming. Future work should also further validate this approach by using an experimental design to compare traditional interviews, retrospective interviews that use only screen captures of programming sessions, and retrospective interviews with screen captures and EDA data. This would allow us to directly measure the added value of EDA data in prompting detailed student recollections of their emotional experiences.

6 LIMITATIONS

While this study contributes new insights and an approach for understanding student emotions, it has a few important limitations. First, we recruited students from a single private university. This limits the generalizability of our findings, because students in a different learning environment might have a different set of emotional triggers or might exhibit different physiological reactions to a programming session. In addition, while our algorithms and methods were designed to reduce the impact of noisy data, there is always the potential for data inaccuracies when using physical sensors, resulting in either false-positives or false-negatives in the EDA peak data. Furthermore, while we incorporated physiological data into our interview protocol with the goal of improving recall of emotions, we are still limited by participants' ability to identify and describe those emotions. Since we provided participants with a list of SCRs during the interview, one concern is that participants may have felt pressure to produce a narrative that explained the peaks. Our data suggests that participants felt comfortable notifying us when no emotion occurred at a given peak, but it is still possible that the presence of the peak data overly influenced student recall of the events that triggered emotions. Conducting a formal study comparing retrospective interviews with and without EDA peak data could help further validate the reliability of this approach for investigating emotional experiences in the future.

7 CONCLUSION

In this paper, we leverage electrodermal activity (EDA) data to prompt student recollections of their emotional experiences while programming during retrospective interviews. This approach allowed us to identify 21 distinct events that trigger positive and negative emotions during programming sessions, providing new information about the experiences that are most salient for students. We also showed that there is a strong relationship between student remembered experiences and the tonic and phasic elements of EDA graphs, demonstrating the expressiveness of EDA data. Our findings suggest that many students experience more negative than positive emotions while programming, and that different students can have very different emotional reactions to the same programming events. This research opens up a number of potential areas for future work, including studies of the relationship between emotions while programming and other factors such as self-efficacy or self-assessments as well as further investigations into the utility of EDA data for prompting student recollections of emotions. We hope that this research inspires a continued focus on the emotional experience of introductory programming students.

ACKNOWLEDGMENTS

We thank our Delta Lab colleagues for valuable feedback. This work was supported by the National Science Foundation under Grant IIS-1755628 and ISS-2045809 and under Grant # 2030859 to the Computing Research Association for the CIFellows Project.

REFERENCES

[1] [n.d.]. *APA Dictionary of Psychology.* American Psychological Association.
[2] [n.d.]. E4 wristband | Real-time physiological signals | Wearable PPG, EDA, Temperature, Motion sensors. https://www.empatica.com/research/e4
[3] Lauri Ahonen, Benjamin Ultan Cowley, Arto Hellas, and Kai Puolamäki. 2018. Biosignals reflect pair-dynamics in collaborative work: EDA and ECG study of pair-programming in a classroom environment. *Scientific Reports* 8, 1 (Dec. 2018), 3138. https://doi.org/10.1038/s41598-018-21518-3
[4] Todd Bentley, Lorraine Johnston, and Karola von Baggo. 2005. Evaluation using cued-recall debrief to elicit information about a user's affective experiences. In *Proceedings of the 17th Australia Conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future.* 1–10.
[5] Paulo Blikstein and Marcelo Worsley. 2016. Multimodal Learning Analytics and Education Data Mining: Using Computational Technologies to Measure Complex Learning Tasks. *Journal of Learning Analytics* 3, 2 (Sept. 2016), 220–238. https://doi.org/10.18608/jla.2016.32.11
[6] Nigel Bosch, Sidney D'Mello, and Caitlin Mills. 2013. What Emotions Do Novices Experience during Their First Computer Programming Learning Session?. In *Artificial Intelligence in Education*, H. Chad Lane, Kalina Yacef, Jack Mostow, and Philip Pavlik (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 11–20.
[7] Nigel Bosch and Sidney D'Mello. 2017. The Affective Experience of Novice Computer Programmers. *International Journal of Artificial Intelligence in Education* 27, 1 (March 2017), 181–206. https://doi.org/10.1007/s40593-015-0069-5

[8] Wolfram Boucsein. 2012. *Electrodermal activity*. Springer Science & Business Media.

[9] Jason J Braithwaite, Derrick G Watson, Robert Jones, and Mickey Rowe. 2013. A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology* 49, 1 (2013), 1017–1034.

[10] Anders Bruun, Effie Lai-Chong Law, Thomas Dyhre Nielsen, and Matthias Heintz. 2021. Do You Feel the Same? On the Robustness of Cued-Recall Debriefing for User Experience Evaluation. *ACM Transactions on Computer-Human Interaction* 28, 4 (Oct. 2021), 1–45. https://doi.org/10.1145/3453479

[11] Yekta Said Can, Niaz Chalabianloo, Deniz Ekiz, and Cem Ersoy. 2019. Continuous Stress Detection Using Wearable Sensors in Real Life: Algorithmic Programming Contest Case Study. *Sensors* 19, 8 (April 2019), 1849. https://doi.org/10.3390/s19081849

[12] Delphine Caruelle, Anders Gustafsson, Poja Shams, and Line Lervik-Olsen. 2019. The use of electrodermal activity (EDA) measurement to understand consumer emotions – A literature review and a call for action. *Journal of Business Research* 104 (Nov. 2019), 146–160. https://doi.org/10.1016/j.jbusres.2019.06.041

[13] Jacqui Chetty. 2013. "I hate programming" and Other Oscillating Emotions Experienced by Novice Students Learning Computer Programming. (2013), 7.

[14] Mi-Hyun Choi, Su-Jeong Lee, Jae-Woong Yang, Ji-Hye Kim, Jin-Seung Choi, Jang-Yeon Park, Jae-Hoon Jun, Gye-Rae Tack, Dae-Woon Lim, and Soon-Cheol Chung. 2010. Changes in Cognitive Performance Due to Three Types of Emotional Tension. In *Database Theory and Application, Bio-Science and Bio-Technology*, Yanchun Zhang, Alfredo Cuzzocrea, Jianhua Ma, Kyo-il Chung, Tughrul Arslan, and Xiaofeng Song (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 258–264.

[15] Juliet Corbin and Anselm Strauss. 2020. *Basics of Qualitative Research (3rd ed.): Techniques and Procedures for Developing Grounded Theory*. Thousand Oaks, California. https://doi.org/10.4135/9781452230153

[16] Mayela Coto, Sonia Mora, Beatriz Grass, and Juan Murillo-Morera. 2021. Emotions and programming learning: systematic mapping. *Computer Science Education* (May 2021), 1–36. https://doi.org/10.1080/08993408.2021.1920816

[17] Hugo D Critchley, Jessica Eccles, and Sarah N Garfinkel. 2013. Interaction between cognition, emotion, and the autonomic nervous system. In *Handbook of clinical neurology*. Vol. 117. Elsevier, 59–77.

[18] James H Cross, Dean Hendrix, and David A Umphress. 2004. JGRASP: an integrated development environment with visualizations for teaching java in CS1, CS2, and beyond. In *34th Annual Frontiers in Education, 2004. FIE 2004*. IEEE Computer Society, 1466–1467.

[19] Michael E Dawson, Anne M Schell, and Diane L Filion. 2017. The electrodermal system. (2017). Publisher: Cambridge University Press.

[20] Ian Drosos, Philip J. Guo, and Chris Parnin. 2017. HappyFace: Identifying and predicting frustrating obstacles for learning programming at scale. In *Visual Languages and Human-Centric Computing (VL/HCC), 2017 IEEE Symposium on*. IEEE, 171–179.

[21] K Anders Ericsson and Herbert A Simon. 1984. *Protocol analysis: Verbal reports as Data*. MIT Press, Cambridge, MA.

[22] Daniela Girardi, Nicole Novielli, Davide Fucci, and Filippo Lanubile. 2020. Recognizing developers' emotions while programming. In *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering*. ACM, Seoul South Korea, 666–677. https://doi.org/10.1145/3377811.3380374

[23] 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. In *Proceedings of the 2020 ACM Conference on International Computing Education Research*. 170–181.

[24] Daniel Graziotin, Xiaofeng Wang, and Pekka Abrahamsson. 2014. Do feelings matter? On the correlation of affects and the self-assessed productivity in software engineering. (2014), 21.

[25] Patricia Haden, Dale Parsons, Krissi Wood, and Joy Gasson. 2017. Student affect in CS1: insights from an easy data collection tool. In *Proceedings of the 17th Koli Calling International Conference on Computing Education Research*. ACM, Koli Finland, 40–49. https://doi.org/10.1145/3141880.3141881

[26] Paivi Kinnunen and Beth Simon. 2010. Experiencing programming assignments in CS1: the emotional toll. In *Proceedings of the Sixth international workshop on Computing education research*. ACM, 77–86.

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

[28] Colleen M. Lewis, Ken Yasuhara, and Ruth E. Anderson. 2011. Deciding to major in computer science: a grounded theory of students' self-assessment of ability. In *Proceedings of the seventh international workshop on Computing education research*. ACM, 3–10.

[29] 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. In *Proceedings of the 2017 ACM Conference on International Computing Education Research - ICER '17*. ACM Press, Tacoma, Washington, USA, 30–38. https://doi.org/10.1145/3105726.3106187

[30] 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. In *In Proceedings of the 2016 ACM Conference on International Computing Education Research*. ACM Press, 211–220. https://doi.org/10.1145/2960310.2960329

[31] Darren Lunn and Simon Harper. 2010. Using Galvanic Skin Response Measures to Identify Areas of Frustration for Older Web 2.0 Users. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A) (W4A '10)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/1805986.1806032 event-place: Raleigh, North Carolina.

[32] 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. In *Proceedings of the 2020 ACM Conference on International Computing Education Research*. ACM, Virtual Event New Zealand, 36–47. https://doi.org/10.1145/3372782.3406279

[33] Cameron McCarthy, Nikhilesh Pradhan, Calum Redpath, and Andy Adler. 2016. Validation of the Empatica E4 wristband. In *2016 IEEE EMBS International Student Conference (ISC)*. IEEE, Ottawa, ON, Canada, 1–4. https://doi.org/10.1109/EMBSISC.2016.7508621

[34] Sebastian C Muller and Thomas Fritz. 2015. Stuck and Frustrated or in Flow and Happy: Sensing Developers' Emotions and Progress. (2015), 12.

[35] Keith Nolan, Aidan Mooney, and Susan Bergin. 2019. An Investigation of Gender Differences in Computer Science Using Physiological, Psychological and Behavioural Metrics. In *Proceedings of the Twenty-First Australasian Computing Education Conference on - ACE '19*. ACM Press, Sydney, NSW, Australia, 47–55. https://doi.org/10.1145/3286960.3286966

[36] Judith S Olson and Wendy A Kellogg. 2014. *Ways of Knowing in HCI*. Vol. 2. Springer.

[37] Martin T Orne. 1962. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist* 17, 11 (1962).

[38] Brian Parkinson and A. S. R. Manstead. 1993. Making sense of emotion in stories and social life. *Cognition & Emotion* 7, 3-4 (May 1993), 295–323. https://doi.org/10.1080/02699939308409191

[39] Ma. Mercedes T. Rodrigo and Ryan S.J.d. Baker. 2009. Coarse-grained detection of student frustration in an introductory programming course. In *Proceedings of the fifth international workshop on Computing education research workshop - ICER '09*. ACM Press, Berkeley, CA, USA, 75. https://doi.org/10.1145/1584322.1584332

[40] Ira J. Roseman. 1996. Appraisal Determinants of Emotions: Constructing a More Accurate and Comprehensive Theory. *Cognition & Emotion* 10, 3 (May 1996), 241–278. https://doi.org/10.1080/026999396380240

[41] Johnny Saldaña. 2021. *The coding manual for qualitative researchers*. sage.

[42] Klaus R. Scherer. 2005. What Are Emotions? And How Can They Be Measured? *Social Science Information* 44, 4 (Dec. 2005), 695–729. https://doi.org/10.1177/0539018405058216

[43] Sara Taylor, Natasha Jaques, Weixuan Chen, Szymon Fedor, Akane Sano, and Rosalind Picard. 2015. Automatic identification of artifacts in electrodermal activity data. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, Milan, 1934–1937. https://doi.org/10.1109/EMBC.2015.7318762

[44] Nanette Veilleux, Rebecca Bates, Cheryl Allendoerfer, Diane Jones, Joyous Crawford, and Tamara Floyd Smith. 2013. The relationship between belonging and ability in computer science. In *Proceeding of the 44th ACM technical symposium on Computer science education*. ACM, 65–70.

[45] Marcelo Worsley and Paulo Blikstein. 2018. A multimodal analysis of making. *International Journal of Artificial Intelligence in Education* 28, 3 (2018), 385–419. Publisher: Springer.

[46] Michal Wrobel. 2018. Applicability of Emotion Recognition and Induction Methods to Study the Behavior of Programmers. *Applied Sciences* 8, 3 (Feb. 2018), 323. https://doi.org/10.3390/app8030323