

# Student Emotions in a Productive Failure Paradigm

by

Zachary Monroe Savelson

A thesis submitted to the Faculty of Graduate and Postdoctoral  
Affairs in partial fulfillment of the requirements for the degree of

Masters of Cognitive Science

in

Masters of Cognitive Science

Carleton University  
Ottawa, Ontario

© 2020, Zachary Monroe Savelson

## **Abstract**

Productive failure is a learning paradigm that reverses the standard order of instruction by asking students to solve problems prior to learning the method to solve them. Despite going against intuitions that instruction should come first, this paradigm has been shown to be effective at engendering conceptual knowledge. To date, however, the role of student emotion in this paradigm has not been investigated. Student emotion influences students' learning outcomes, and while there are some exceptions, negative emotion typically interferes with student learning. This leads to a conundrum because productive failure has been shown to have a positive effect on learning, whereas the negative emotion that might be experienced while failing in the first stage of productive failure could have a detrimental effect. To address this, we investigate the presence of emotions in productive failure as well as their changes over time and their relations to learning outcomes. We found that productive failure elicits similar emotions as standard learning paradigms and that instructional activity has a significant effect on emotion. We did not however, find evidence for relations between emotions and learning.

## **Acknowledgements**

Thank you first and foremost to my friends and family for their constant support and guidance and especially to my mother, who pushes me to always be at my greatest. A sincere thank you goes to Dr. Kasia Muldner whose knowledge, infinite patience and guidance made this project possible in the first place. I would also like to thank the Institute of Cognitive Science for being the most welcoming and excellent faculty I have had the pleasure to work with.

# Table of Contents

<b>Abstract.....</b>	<b>i</b>
<b>Acknowledgements .....</b>	<b>ii</b>
<b>Table of Contents .....</b>	<b>iii</b>
<b>List of Tables .....</b>	<b>vi</b>
<b>List of Figures.....</b>	<b>vii</b>
<b>Chapter 1: Introduction .....</b>	<b>1</b>
<b>Chapter 2: Related Work.....</b>	<b>4</b>
2.1    Learning from Failure.....	4
2.2    Productive Failure .....	8
2.2.1    Productive failure paradigm.....	8
2.2.1.1    Phase 1: exploration and generation.....	8
2.2.1.2    Phase 2: consolidation. ....	10
2.2.2    Proposed mechanisms responsible for promoting learning.....	10
2.2.3    Productive failure in practice .....	11
2.3    Emotions and how they Impact Learning.....	14
<b>Chapter 3: Methodology.....</b>	<b>20</b>
3.1    Participants .....	20
3.2    Materials: Instructional Materials.....	21
3.2.1    SNAP! .....	21
3.2.2    The productive failure problem: sorting task .....	22
3.2.3    Instructional videos .....	24
3.2.3.1    Tutorial video .....	24
3.2.3.2    Sorting lesson .....	26

3.2.4	Cheat sheets.....	27
3.2.5	Pretest and posttest.....	28
3.3	Materials: Questionnaires.....	29
3.3.1	Demographic questionnaire.....	29
3.3.2	The academic emotions questionnaire (AEQ).....	29
3.3.3	The self-efficacy questionnaire.....	30
3.3.4	The curiosity questionnaire.....	30
3.3.5	The intolerance for uncertainty scale.....	30
3.3.6	The experience sampling questionnaire.....	31
3.4	Design and Procedure.....	31
3.5	Qualitative Coding.....	34
3.5.1	Coding Scheme: Details.....	36
<b>Chapter 4: Results.....</b>		<b>39</b>
4.1	Emotions Expressed During Productive Failure.....	39
4.1.1	Examples of emotions expressed during productive failure.....	40
4.2	State and Trait Emotions.....	46
4.3	Learning and Emotions.....	50
4.3.1	Relationships between emotions and learning.....	52
<b>Chapter 5: Discussion.....</b>		<b>55</b>
5.1	Emotions and Related States in Productive Failure.....	55
5.2	Learning and Emotion in Productive Failure.....	59
5.3	Limitations and Future Work.....	64
5.4	Conclusion.....	67
<b>References.....</b>		<b>68</b>
<b>Appendices.....</b>		<b>78</b>

Appendix A .....	78
A.1 Pretest.....	78
A.2 Posttest .....	79
A.3 Grading Scheme .....	82
Appendix B.....	83
B.1 SNAP! Interactive Cheat Sheet.....	83
B.2 SNAP! Block Definitions and FAQ Cheat Sheet.....	84
Appendix C.....	87
C.1 Demographic Questionnaire.....	87
C.2 Academic Emotions Questionnaire (AEQ).....	88
C.3 Self-Efficacy Questionnaire .....	92
C.4 The Curiosity Questionnaire .....	93
C.5 The Intolerance for Uncertainty Scale .....	94
C.6 Experience Sampling Questionnaire .....	95
Appendix D .....	96
D.1 Skew and Kurtosis Values for Experience Sampling Measures .....	96
D.2 Correlations Between Academic Emotions and Total Posttest Score, postTotal, Controlling for Pretest.....	97
Appendix E.....	98

## List of Tables

Table 4.1: <i>Mean and standard deviation of expressions related to emotions, failure, and ideas</i> .....	40
Table 4.2: <i>Mean and standard deviation for the AEQ trait academic emotions (/5)</i> .....	47
Table 4.3: <i>Mean and standard deviation for the other trait academic emotions</i> .....	47
Table 4.4: <i>Mean and standard deviation for the experience sampling data (shown as mean followed by standard deviation in brackets)</i> .....	48
Table 4.5: <i>Mean and standard deviation for the pretest and posttest scores (shown as raw scores followed by percentages in brackets)</i> .....	51

## List of Figures

Figure 3.1. SNAP! interface with a sample program.....	22
Figure 3.2. SNAP! sorting template that includes custom blocks (left) and list creation program (middle) .....	24
Figure 3.3. Tutorial Program .....	26
Figure 3.4. Worked example of sorting .....	27
Figure 4.1. Mean number of utterances for confusion, frustration, uncertainty, and failure, per participant. Ideas were not included as they occurred more and affected the scale of the graph (error bars represent 95% confidence intervals) .....	40
Figure 4.2. AEQ trait academic emotions (error bars represent 95% confidence intervals) .....	47
Figure 4.3. Change in emotion (possible range 1 – 5) over the three target timepoints for each of the five target emotions (error bars not shown because they made the figure unreadable – see Table 4.4 for the variance information for these variables) .....	48
Figure 4.4. Scores for pretest, as well as the procedural section, recall section, conceptual section, and postTotal (sum of scores) (error bars represent 95% confidence intervals) .	52

## Chapter 1: Introduction

Various theories of impasse-driven learning show that failure is an important part of learning. Failure can have both negative and positive effects. On the one hand, it can result in students feeling discouraged and frustrated, but on the other hand, it can promote reflection and serve as a catalyst for new avenues of reasoning to overcome knowledge gaps. The present thesis focuses on a particular failure-centered paradigm called *productive failure*. This paradigm manipulates when students experience failure and is split into two phases: exploration/generation and consolidation. In the first phase, students are asked to solve problems that they do not have the adequate background to solve, which leads them to fail at finding the correct solution. In the subsequent consolidation phase, instruction is provided both on how to solve the problem and the pertinent concepts behind it. This ordering may seem counterintuitive as it is contrary to traditional instruction where students solve problems only after having learned how to solve them. However, there are benefits to productive failure: students learn more from this paradigm than traditional instruction, as established by numerous studies (Kapur, 2008; 2012; Kapur & Kinzer, 2009; Loibl & Rummel, 2014). What has yet to be investigated is the role of student emotion in productive failure.

Research outside of the productive failure paradigm has shown that students' emotions impact their learning (D'Mello et al., 2014; Kim & Pekrun, 2014; MacIntyre & Vincze, 2017). Positively valenced emotions generally improve learning while, with some caveats, negatively valenced emotions decrease it. For example, when students take pride in their work and feel hopeful that they can succeed, they have better learning outcomes (Kim & Pekrun, 2014). Conversely, negative emotions such as boredom and

anxiety have consistently been associated with reduced learning (Craig et al., 2004; Pekrun, Elliot, & Maier, 2009). The results related to frustration have been mixed: sometimes the relation between frustration and learning is negative (Kort, Reilly & Picard, 2001) and sometimes it is positive, dependent on the level of frustration (Pataranutaporn, & Liu, 2013). In addition to these basic emotions having an effect on learning, researchers have also found that related states such as self-efficacy, curiosity, and intolerance for uncertainty also influence learning (Buhr & Dugas, 2002; Litman & Spielberger, 2003; Imam, 2007).

To date, student emotion has not been analyzed in the productive failure paradigm. In other paradigms, failure has been shown to elicit negative emotions such as anxiety, frustration, confusion, and, in some cases, hopelessness (Nummenmaa & Niemi, 2004). Thus, it seems counterintuitive that productive failure improves learning, given that it also elicits failure. To investigate this conundrum, we examined the presence and change of emotions over time in a productive failure paradigm as well as their impact on learning. To our knowledge, this is the first study to investigate emotion during productive failure and because of this, we took an exploratory approach. Specifically, we measured emotion both through standard questionnaires, as well as through experience sampling and protocol analysis based on transcripts of students working in pairs on a productive failure problem-solving task. These sessions were recorded and analyzed for the presence of emotions during productive failure. We also analyzed how emotions changed based on the instructional task, and their relationship with learning from productive failure.

We hypothesized that, like in other learning paradigms, emotions will be present in productive failure; specifically, negative emotions such as anxiety, confusion, and frustration. Based on prior work, we also hypothesized that instructional activity would impact the emotions measured.

In addition to being the first study of emotion in the productive failure paradigm, this work is also the first to use a novel task/domain, namely computer programming. Programming is something that many students struggle to grasp (Costa & Miranda, 2017) and so is well suited to a task that requires students to fail. We use a particular programming task (sorting a list) because completing this task requires knowledge of several concepts. In addition to the novel domain used, no study to date has attempted to examine productive failure's outcomes in an older (university level) population.

## **Chapter 2: Related Work**

### **2.1 Learning from Failure**

This section describes work on the impact of failure during instructional activities in paradigms other than productive failure, while Section 2.2 describes productive failure. In academic contexts, failure is defined as not reaching the solution to a problem. While it may seem counterintuitive, failure can increase student learning (Loibl & Rummel 2014; Kapur, 2016; Darabi et al., 2018), although this only occurs under certain conditions. For example, Falout et al. (2009) found that when students were given no guidance and failed because the task was beyond their capabilities, they learned considerably less than their peers who were not exposed to failure. However, when failing students are able to activate prior knowledge (Schank, 1999), become aware of their own knowledge gaps (VanLehn, 1998), and recognize deeper features of the problem (Tawfik, Rong & Choi, 2015), they develop a deeper understanding of the problem space that leads to increased learning. Researchers have proposed different theories for why this increase in learning occurs.

Piaget (1977) proposed the theory of cognitive disequilibrium. In this theory, learning has two parts: assimilation (incorporating new information into existing schemas) and accommodation (adaptation of one's schema to reflect reality). Piaget suggests that perturbations of cognitive schemas can occur when a person encounters information or a problem that does not fit an existing schema. He posits that these perturbations are critical for learning because they help learners realize that their schemas are insufficient to process the incoming new information. Failure is a prime example of a

perturbation, as it causes people to reexamine the state of their world and to adapt their schema of that space accordingly.

Leveraging Piaget's theory, VanLehn (1998) suggested that impasses encountered while problem solving are critical for learning. Impasses are temporary or permanent failures to find a problem solution. For example, when solving an algebra equation, if a student cannot keep simplifying the equation, that is an impasse. When students encounter impasses, they can aim to overcome them in a number of ways, such as by learning new knowledge from teachers, textbooks, or from comparing other solutions. The new knowledge can then be incorporated into their existing schema.

Another theory on learning from failure is the Failure-Driven Memory framework. This framework was proposed by Roger Schank (1999), who focused on how people develop mental scripts for sets of circumstances. A script is a robust memory structure consisting of the knowledge necessary to solve a given problem. When a novel problem is encountered and a relevant script is not present, the person experiences failure. This makes them investigate the causes of the failure and generate explanations. If those explanations are valid, they are added to the learner's schema of the problem space. (Tawfik, Rong, & Choi, 2015). This theory operates on a similar principle as VanLehn's: encountering failure or an impasse leads learners to confront their errors, learn from them, and use this new knowledge to update their schema of the problem/solution space.

Based on theories of learning from failure, Tawfik, Rong and Choi (2015) developed a comprehensive model of learning from failure, which involves a number of stages. The first stage involves defining the conditions for success. Once defined, learners

can use their current mental model to explore the problem, implement solutions, and investigate the viability of the outcome. If the outcome is positive, then the mental model for the problem is rehearsed for future problem solving. If the outcome is negative, the learner experiences failure and challenges their mental model. This is done by asking why the solution failed and then validating those reasons. Once the reasons are known, learners can identify the root cause of their failure and use that to generate, implement, and compose a new solution. If the novel solution is successful, their mental model is updated. If the novel solution is still not viable, the cycle is repeated. This model is useful because it highlights the critical components for learning from failure (listed above) and reinforces their importance in failure-based learning.

Tawfik et al. (2015) propose four guidelines for designing ‘learning from failure’ paradigms. The first guideline states that conditions for failure must be defined, i.e., learners should be explicitly aware of the conditions for both success and failure in a problem space. The second guideline proposes that learning environments should be designed so that students encounter failure. The third guideline involves supporting student inquiry into why the failure occurred. Students can use this information to update their mental schemas of the problem space and to reflect on their failures and misconceptions (Piaget, 1977). To accomplish this, students can be asked to explicitly generate and justify reasons for their failure (Schank, 1999). In practice, students can be asked to review their model in comparison to a correct one and discuss what went wrong with their model (Schank, 1999; Kapur, 2011). The fourth guideline centers around providing tools to help students overcome the failure.

Aligned with the work above, Gadgil, Nokes-Malach, and Chi (2012) conjectured that it is the confrontation of one's errors that leads to increased learning. To test their claim, they asked students to either compare their own failed solution to an expert one or to self-explain the expert solution aloud. They found that when students confronted their own errors by comparing failed and expert solutions, their knowledge of the topic increased more than if they had just self-explained an expert model that did not contain errors.

Sinha & Kapur (2019) and Loibl & Rummel (2014) also showed the importance of confronting one's errors: Both studies found that when errors were confronted after problem solving, students learned more. These results suggest that confronting errors after problem-solving is important for learning from failure. A consideration, however, is who made the error. Kapur (2014) showed that examining errors made by other people is not as effective as failing and examining one's own errors. While confronting one's errors is beneficial for increasing the effectiveness of 'learning from failure' paradigms, this may be difficult to realize in practice because it would be difficult for every error made in a large group of students to be recognized and addressed by the instructor.

There is some evidence that the relationship between failure and learning is curvilinear: too little or too much failure can be detrimental, while the "right" amount increases learning (He et al., 2017). The relationship works this way because of what happens when one fails. Failing leads to the confrontation of one's errors and incorporating what is learned from that into future solutions. If students do not fail 'enough', they will not confront their errors and will learn less overall. On the other hand, if students fail too much, they will become frustrated and learning will be equally

stymied. The ‘goldilocks’ zone of failure is reached by allowing students to fail enough so that they make errors but not so much that they become frustrated at their lack of progress.

## **2.2 Productive Failure**

In the previous section, we described research on failure paradigms that do not address whether *when* the failure takes place has an effect on student learning. This latter aspect was investigated by Manu Kapur and colleagues (2008; 2012; 2016) using the productive failure paradigm. Most of these studies involve the domain of mathematics (Kapur, 2008; 2012) but there is some work that examines productive failure in other domains like physics and biology (Kapur & Kinzer, 2009; Toh & Kapur, 2017).

### **2.2.1 Productive failure paradigm**

There are two phases in the productive failure paradigm: (1) problem exploration and solution generation; (2) instruction and consolidation. Productive failure involves having students first explore and generate various potential problem solutions and only after be given instruction on the canonical solution. Below is a description of each phase.

#### **2.2.1.1 Phase 1: exploration and generation.**

The goal of the first productive failure phase is for students to explore the problem space and generate multiple solutions - whether these are correct is not relevant (Kapur & Bielaczyc, 2012). To ensure adequate exploration and generation, there are four criteria for the problem space that must be met: (1) The problems must be complex in nature; (2) The problem space must allow for multiple solution methods to be generatable; (3) The problems must be challenging in nature but not so challenging as to

induce frustration; (4) The learner should use prior knowledge to attempt to solve the problem (Kapur, 2016).

We now briefly expand on these requirements. Complex problems refer to problems without obvious solutions and/or those that have multiple paths to the same solution. They also have several underlying concepts to them and are not ‘one-dimensional’ (Kapur & Bielaczyc, 2012). To illustrate, consider a simple algebra problem that asks for the solution of a math equation. There is one solution to the problem and only one way to approach solving it. An example of a complex problem, commonly used in productive failure studies, involves calculating the variance in a set of numbers without knowing the formal definition for variance. There are many concepts behind the problem and multiple methods to approach solving it. As far as the third requirement, namely challenge, if the problem is not challenging enough, students will not explore the problem space. Without exploration, they cannot become aware of their knowledge gaps and/or activate their prior knowledge. If the problem is too challenging, Kapur hypothesizes that frustration would set in and students will not learn (Kapur, 2016; He et al., 2017). As far as the fourth requirement, some prior knowledge is necessary for productive failure because it gives students grounding within the novel problem space (Toh & Kapur, 2017). Importantly, Kapur and colleagues (2012; 2016) make it clear that no guidance is to be given to students during phase 1; students are to rely only on their prior knowledge and intuition in their attempts to find a solution (Kapur & Bielaczyc, 2012).

The final design element of the first phase is that the exploration is done collaboratively. Kapur & Bielaczyc (2012) discuss how collaboration has been shown to foster increased sharing, elaboration, and exploration of the problem space.

Consequently, productive failure studies typically involve students working collaboratively during phase 1.

### **2.2.1.2 Phase 2: consolidation.**

After phase 1, productive failure moves on to the instruction/consolidation phase. This phase typically consists of a group discussion led by a teacher describing the affordances and constraints of various solution methods, including the canonical solution as well as failed solutions. To show the differences between failed and canonical solutions, the instructor can introduce contrasting cases. A contrasting case varies one element that makes the solution incorrect, showing students the importance of specific elements of the canonical solution. Thus, contrasting cases show a failed solution and the reason why it failed. Discussing failed solutions allows students to compare them with the canonical one, making explicit the knowledge gaps that led to their failure; they also become aware of the deeper conceptual features of the problem space (Kapur & Bielaczyc, 2012). Kapur and Bielaczyc also state that the instructor should explicitly encourage participation in the activity from all learners.

### **2.2.2 Proposed mechanisms responsible for promoting learning**

Productive failure is meant to activate four key mechanisms that promote learning: (1) activation of prior knowledge related to the target concepts; (2) attention to the critical conceptual features of the target concepts; (3) explanation and elaboration of those features; (4) organization and assembling of the features into fully formed target concepts (Kapur & Bielaczyc, 2012).

While productive failure studies have not validated the effectiveness of these individual mechanisms, productive failure as a whole has been shown to increase

learning. However, there is a caveat: productive failure increases conceptual knowledge while having no significant effect on procedural knowledge (Kapur, 2008; Kapur & Bielaczyc, 2012; Loibl & Rummel, 2014). Procedural knowledge is ‘how to’ knowledge, or the knowledge of the procedure for accomplishing a goal. In problem-solving contexts, it can be defined as the ability to execute a series of actions in order to solve a problem (Rittle-Johnson, Siegler & Alibali, 2001). Conceptual knowledge is the ‘why/what’ knowledge and so deals with the concepts that underly the problem and its solution (Rittle-Johnson, Siegler, & Alibali, 2001). To illustrate using the concept of variance (Kapur, 2008), the procedural knowledge involves computing the answer to the problem by plugging numbers into the formula, while the conceptual knowledge involves understanding that variance corresponds to the spread of numbers from the mean. Productive failure only impacts conceptual knowledge because it is designed to expose students to the conceptual problem features as will be described below – while students are also exposed to the procedure for solving the problem, this is not the focus.

### **2.2.3 Productive failure in practice**

As mentioned above, productive failure increases conceptual knowledge. To illustrate, Kapur (2012) evaluated productive failure in a classroom with ninth-grade students learning about the concept of variance. There were two study conditions: productive failure and *direct instruction*. Each condition (productive failure and direct instruction) involved four class periods. For the productive failure group, the first two periods were spent having students work in triads to solve a variance problem dealing with the level of consistency displayed by soccer players. Because students had no prior instruction on the concept of variance and no guidance, they had to rely on their general

prior knowledge about soccer to attempt to generate a quantitative index for consistency. The third period was when the instruction took place. The instructor compared and contrasted different solution methods (both failed and canonical) and explained the concept of variance.

For the direct instruction condition, the order of instruction was flipped: the first period consisted of the teacher explaining the concept of variance. This was done by having students do worked examples followed by problem solving. After solving the isomorphic problem, the errors, misconceptions, and conceptual aspects of the problem were discussed class wide. Note that contrasting cases were not used in the direct instruction condition. This means that the instructional periods in the two conditions were different, which is a potential source of bias. The second and third periods both consisted of problem solving. After a problem was solved, the teacher discussed the solutions. The fourth period for both conditions was the same: students solved three test problems. The day after the intervention, students were given a posttest that consisted of procedural, conceptual, and transfer items (taking what was learned and applying it to a novel context). Kapur found a significant difference between performance on both the conceptual and transfer measurements in favor of productive failure over direct instruction. As expected, there was no difference between productive failure and direct instruction on the procedural measure. This shows that conceptual learning and ability to transfer to novel contexts are both improved by productive failure more than direct instruction. This result has been validated by multiple other studies (Kapur, 2008; Kapur & Kinzer, 2009; Kapur & Bielaczyc, 2012).

As alluded to above, however, there is a limitation to the traditional productive failure studies: the instructional phases for productive failure and direct instruction are different because direct instruction does not include exposure to contrasting cases. This is a confounding variable because the presence of contrasting cases could be driving the learning as opposed to instructional order. To address this limitation, Loibl & Rummel (2014) created a productive failure paradigm (Exp. 2). Specifically, they used a 2x2 experimental design manipulating the order of instruction (before or after problem solving, with before corresponding to productive failure and after corresponding to standard instruction) and the form of the instruction (including or not including contrasting cases).

The problem the students had to solve dealt with mathematical variance, a problem from Kapur (2012). Students who were exposed to this ‘failure paradigm’ outperformed their peers who were exposed to regular instruction with respect to learning. As well, the main effect of form of instruction was significant, favoring instruction based on contrasting cases. These results show that productive failure is an effective paradigm and also that the presence of contrasting cases increases learning.

Productive failure was designed as a collaborative paradigm, typically done with dyads or triads (Kapur, 2012; Kapur & Bielaczyc, 2012). Mazziotti et al. (2015) investigated whether the collaborative nature of the productive failure phase had an impact on the paradigm’s outcomes. To do this, they ran two quasi-experimental studies varying collaboration firstly in just the productive failure condition (experiment 1) and then in both conditions (experiment 2). The first study had fourth grade students complete a productive failure activity either individually or together; there was also a collaborative

direct instruction condition. They found no significant difference between collaborative and individual productive failure learning but both productive failure conditions (collaborative and individual) showed greater learning than the direct instruction condition. The second study was a 2x2 design varying both instruction timing (productive failure vs direct instruction) and social structure (collaborative vs individual). They found that individual productive failure learners showed more improvement than collaborative ones. However, this one result contradicts many prior studies showing that collaboration is beneficial for learning (Kapur & Bielaczyc, 2012). A potential explanation for this discrepancy is that the population was young (elementary students) and may not have known how to effectively collaborate. We decided to follow productive failure's original design and thus have students collaborate during phase 1, as we wanted to investigate the impact of emotion in the original paradigm. As such, in our study students worked in pairs for the productive failure phase.

### **2.3 Emotions and how they Impact Learning**

The last two sections described the role of failure in learning, including the timing of when failure occurs. Another factor relevant to the current research is how students feel during learning activities, i.e., student emotion (Pekrun & Stephens, 2010; Kim & Pekrun, 2014; MacIntyre & Vincze, 2017). Emotion can be defined as: "... an episode of interrelated, synchronized changes ... in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism" (Scherer, 2005). In other words, emotions are physiological and psychological responses to stimuli in our environment. Scherer characterizes emotions as having the following components: (1) emotions must be the result of a specific event/stimulus, (2) emotions must be appraisal

driven, (3) the response to the emotion should match that appraisal, (4) emotional states are prone to rapid change as appraisals change, (5) emotional responses have a specific behavioral impact, (6) Emotions are more intense than moods and (7) emotions last less time than moods. Emotions are distinct from feelings, in that feelings are the psychological part of the overall experience of an emotion (Scherer, 2005), while emotions also have a physiological component. As noted above, emotions are also distinct from moods. Moods are long-lasting states that generally have low levels of arousal whereas emotions are usually shorter bursts of arousal (low or high).

Emotion can be characterized using the dimensions of valence and arousal (Wundt, 1905). Valence describes whether the emotion is a positive ('good') or negative ('bad') emotion. For example, happiness is a positively-valenced emotion while anger is a negatively-valenced emotion. Arousal corresponds to the physical and/or mental change in activity in response to an emotional stimulus; it has also been characterized as the level of activation an emotion produces. For example, annoyance produces medium arousal while outrage produces high arousal.

Emotions affect various aspects of student learning, including cognitive processes and strategies, decision-making, and motivation (Kim & Pekrun, 2014). These are bi-directional relationships: not only does emotional state affect learning, but what and how students learn affects their emotions. Cognitive processes can be initiated, altered, accelerated, and stopped by emotions. These effects depend on the emotional components of valence and arousal (Pekrun et al., 2002; Craig et al., 2004; MacIntyre & Vincze, 2017). To illustrate, MacIntyre & Vincze (2017) showed how valence affects learning by studying the emotional states of 182 Italian secondary students learning English Second

Language (ESL). Their survey measured several constructs related to second language learning, which were then correlated with a set of 10 positive and 9 negative emotions. They found that all but four of the positive emotions were strongly and positively correlated with L2 learning variables and, while negative emotions did correlate negatively with learning, the trends were less strong and less consistent. Another example of valence impacting learning is that students in positively-valenced emotional states use more general knowledge and heuristics whereas people in negative states use more systematic, detail-focused analyses (Kim & Pekrun, 2014). In a similar vein, people make decisions that maximize positive emotions and minimize negative emotions (Kim & Pekrun, 2014). For example, students who failed at solving a problem may be less motivated to attempt the problem again out of fear of experiencing the negative emotions that came with failing. Moreover, motivation changes based on the valence of the emotions associated with that task (Kim & Pekrun, 2014). For instance, people are less motivated to do tasks with which they associate a negative emotion and vice versa.

For arousal, the effects are different depending on the valence of the emotion (Pekrun et al., 2002; Pekrun, 2006). For positive emotions, the higher the arousal, the greater the increase in learning and vice versa. For negative emotions, low arousal emotions lead to decreases in learning whereas high arousal emotions can, under certain circumstances, increase learning while under other circumstances, decrease learning.

The present work focuses on achievement emotions (Pekrun, Elliot & Maier, 2009; Pekrun & Stephens, 2010; Pekrun et al., 2011). These are the emotions that relate to activities or outcomes and have been shown to impact student learning (Pekrun, Elliot & Maier, 2009; Villavicencio & Bernardo, 2013; Goetz et al., 2015). Positive, activating

achievement emotions (i.e. enjoyment, hope, and pride) have a positive effect on learning (Pekrun et al., 2002; Villavicencio & Bernardo, 2013). For negative emotions, deactivating emotions related to prospective outcomes (e.g. hopelessness) have been found to consistently decrease learning. The relationships with activating negative emotions are more complex. While anxiety, anger, shame, boredom, and frustration typically correlate negatively with performance (Pekrun et al., 2002; Goetz & Hall, 2013), some studies have found the opposite. For example, Lane et al. (2005) found that anger felt prior to an exam interacted with the depressed mood of participants and led to increased performance in participants who did not report depressive mood symptoms. Thus, both the arousal and valence of an emotion affect learning outcomes - while negatively valenced emotions are usually detrimental, there are some instances where they can benefit learning.

Some of the studies above used questionnaires to measure emotions, typically once during the experiment. Other methods are designed to measure emotions more frequently. In Baker et al. (2010), in one experiment, students self-reported emotions retrospectively after watching a video of themselves working on the problems; in a second experiment, trained judges rated students' expressions of emotion at specific timepoints. In Muldner et al. (2015), the experience sampling method was used, where participants were asked to self-report emotions at several timepoints using a Likert scale.

Aside from basic emotions like enjoyment, anger and boredom, there are more complex emotions that affect learning. Confusion is one such state. D'Mello et al. (2014) hypothesized that confusion might be beneficial to learning because it leads to impasses and error confrontation. The authors state that confusion "arises out of information-

oriented appraisals of the extent to which incoming information aligns with existing knowledge structures and whether there are inconsistencies and other discrepancies in the information stream.” (D’Mello et al., 2014, p. 154). In their study, confusion was introduced by having participants interact with animated computer agents. One of the agents would introduce an incorrect and/or contradictory opinion that contrasted with the opinion of the other agent and then, the participant had to choose which opinion had more scientific merit. There was a significant increase in participants’ posttest performance when they became confused during the learning phase. The authors proposed that becoming confused allows students to go through the same mental process as during an impasse; namely, they fail, they confront their failure and in doing so, access prior knowledge. In other words, becoming confused promotes the same cycle as outright failure: failing, confronting and recognizing one’s errors, and then learning from those errors.

Beyond the emotions described above, there are also related states that impact learning – ones relevant to the present research include self-efficacy, curiosity, and uncertainty. Self-efficacy is a measure of one’s own belief in ability to master a task and succeed. To show how this construct is related to learning, Sherer et al. (1982) created the self-efficacy scale that measured both general and social self-efficacy. They found that increased self-efficacy was positively related to learning gains, i.e. the more confidence an individual has in their ability to succeed, the more they will learn.

As far as curiosity, Oudeyer et al. (2016) found that there is a positive relation between curiosity and learning and that this relation is bi-directional. They suggest that

when learning occurs, there is an increase in the desire to learn more, which in turn leads to increased curiosity.

Uncertainty had also been positively related to learning. Lamnina and Chase (2019) found that when students were exposed to uncertainty (whether expected or unexpected), they were more curious than those students who were not exposed to uncertainty. They also found that students in the uncertainty conditions were better able to transfer the knowledge they learned into novel contexts. Other studies have found that how students feel about uncertainty also has an impact. For instance, tolerance for uncertainty correlates positively with L2 (second language) learning and metacognitive ability (Chapelle & Roberts, 1986; Keshavarz & Assar, 2009).

The work described above shows that both failure and emotions have distinct effects on student learning. Failure can be beneficial or hindering depending on the context, and emotional states' influence varies based on the valence and arousal level of the emotion(s). The goal of the present study was to examine the presence of emotions in a productive failure paradigm and how they changed after each instructional task. Our second goal was to investigate whether those emotions related to learning.

## Chapter 3: Methodology

### 3.1 Participants

The study participants were 50 university students (23 males, 22 females, 5 declined to answer) between the ages of 17 to 29 ( $M = 21$ ,  $SD = 3.2$ ) recruited through word of mouth, Facebook posts, and Carleton University's SONA participant system. In order to participate, individuals had to come to the study with a friend (as the study involved working with one other individual). Both participants had to have some self-reported knowledge of computer programming, either from tinkering, from a high school course, or from one university-level computer science course. To avoid ceiling effects, participants who had taken more than one university programming class were not eligible. All participants provided informed consent. When recruiting participants, we asked them if they were currently taking a course with the thesis supervisor Dr. Muldner. If they were, they were given a modified consent form that informed them that participating in this experiment would have no impact on their grade for that class and that Dr. Muldner would not be aware of their participation. This was done in order to avoid a potential conflict of interest. Participants received either 2% course credit for CGSC 1001 or 25\$ (those enrolled in CGSC 1001 could choose either option). One pair of participants was unable to complete the study due to internet connection issues that caused their session to end prematurely. Thus, only 48 participants were considered in the analysis (some of the analyses involved 46 participants due to data loss; when this happens, it is clearly stated in the thesis).

## **3.2 Materials: Instructional Materials**

### **3.2.1 SNAP!**

We used the web-based SNAP! programming environment in the experiment. SNAP! was created by the University of California at Berkley (Berkley University, (N/A) Snap!) to help students learn to code (See Figure 3.1) and indeed there is evidence that it is especially useful for novice programmers (Costa & Miranda, 2017). One of the reasons for this is that SNAP! abstracts away the need to learn programming language syntax. Specifically, instead of writing programs in a language like Python, users build programs by snapping blocks together, where each block has its own built-in syntax. For example, instead of having to write the syntax for a loop, users select a ‘Repeat’ block in the SNAP! interface and drag it over to the coding area. This design allows participants to focus on code semantics without having to deal with syntax. This was important considering that we recruited participants with limited programming knowledge.

The SNAP! interface includes many different types of blocks corresponding to the various programming constructs. For the study, we customized the SNAP! interface by creating a template so that the only blocks participants were able to see/use were those directly relevant to the task at hand. This was done in order to avoid participants becoming overwhelmed by the amount of blocks present in the original interface and to allow them to focus on only those blocks that were necessary for the task.

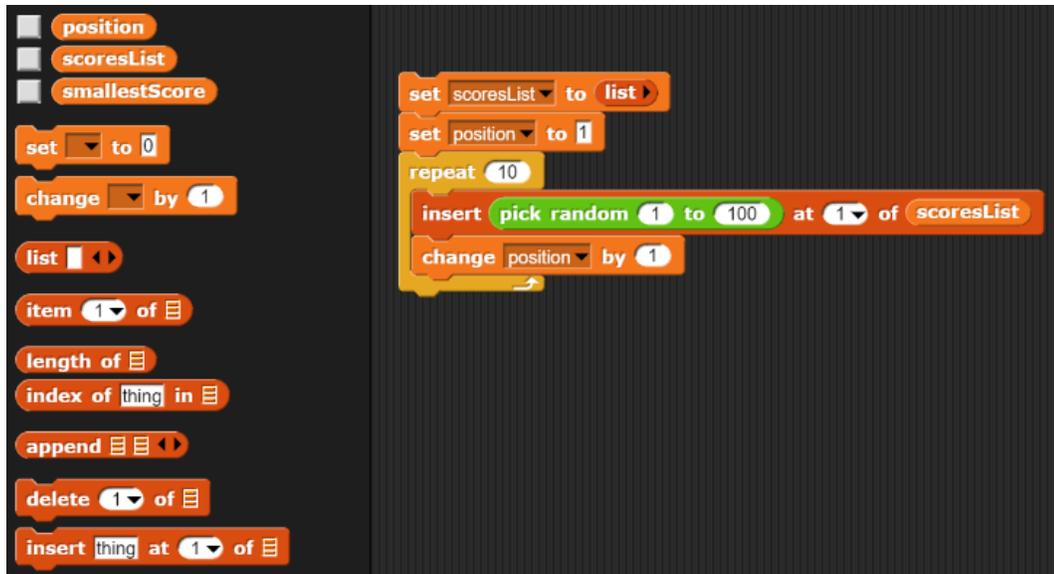


Figure 3.1. SNAP! interface with a sample program

### 3.2.2 The productive failure problem: sorting task

In order to apply the productive failure paradigm to a programming context, we had to devise an appropriate programming problem. The problem had to be complex enough to satisfy productive failure’s constraints (see Section 2.2 for details), be open ended to afford exploration of multiple solutions, and also be technically completable<sup>1</sup> in the allotted time span. Sorting a list of numbers meets these criteria.

There are many algorithms for sorting a list. For this experiment, we used a version of the *selection sort* algorithm. This algorithm requires understanding of four concepts (that we labelled as ‘big ideas’): (1) A nested loop is needed to sort a list, including an inner loop responsible for finding the position of smallest element of the list and an outer loop responsible for inserting that element into the sorted list. (2) The inner

---

<sup>1</sup> Technically completable means that the problem should be solvable within the time limit for experienced individuals. In the present context, we specifically recruited participants unlikely to reach a solution because this is a pre-requisite of the productive failure paradigm.

loop needs to find the *position* of the smallest element. Position is required because removing an item from one list and inserting it into another in SNAP! requires information on that element's position rather than its value. (3) A secondary list is used to hold the sorted values (we recognize there are methods for sorting with one list, but they add complexity to the solution). If this secondary list is not included and the values are simply appended to the end of the original list, the part of the code that finds the smallest element would consistently pick the value that was appended (the smallest one) and would never consider any other values. An additional problem with not having a second list is that the loop responsible for inserting values into the list would never stop running because it is designed to continue as long as there are values in the list (both of the previous two issues could be avoided by stopping the loop early through additional code but that adds complexity to the solution). (4) Once an element is inserted into the new (sorted) list, it must be removed from the original list. As mentioned above, this is needed so that the code that finds the smallest element does not always pick the first smallest element.

In addition to the conceptual knowledge required to implement the sorting algorithm (i.e., the 'big ideas' presented above), some of the procedural aspects involved in sorting a series of values include implementing loops and conditional statements in SNAP!, creating and using variables to store values, and using lists.

The sorting task was realized with SNAP! using a custom template that came preloaded with a snippet of code to create a list of random numbers between 1-100 (see Figure 3.2). The code was identical to the one shown in the tutorial video described below and initialized the list that was eventually to be sorted – it was provided because

we were only interested in participants' ability to implement a sorting algorithm and did not want them to waste time re-creating a list initialization procedure.

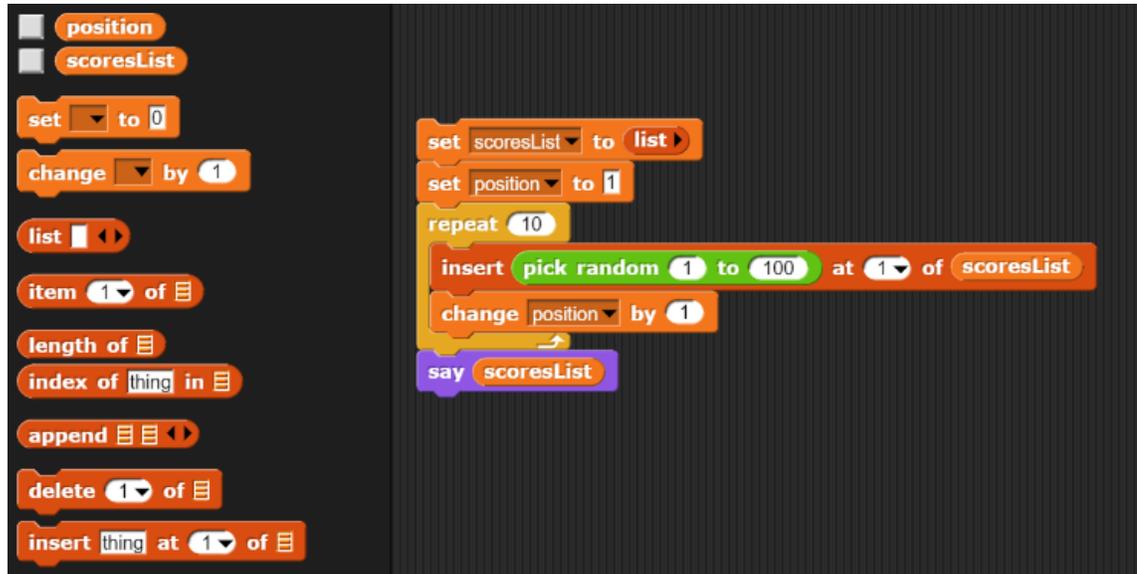


Figure 3.2. SNAP! sorting template that includes custom blocks (left) and list creation program (middle)

### 3.2.3 Instructional videos

#### 3.2.3.1 Tutorial video

We developed a tutorial video designed to provide background on key programming constructs and on how to implement them in SNAP! This video is distinct from the lesson video viewed after the productive failure (the latter is described below). The tutorial video consisted of screen capture of a SNAP! program being created (i.e. blocks being snapped together) with a narrator (the thesis author) describing what was being done and why. The total length of the video was 19 minutes and 41 seconds. To design the video, we first identified the prior knowledge that sorting requires, and then created a script that described what was happening on-screen as well as commentary

concerning the actions the narrator was taking (the script helped to maintain clarity of presentation for the final tutorial video).

The tutorial video described how to create a program that prints out the smallest score in a list of numbers. This topic teaches participants about some programming prerequisites for sorting such as variables, loops, and conditionals. While some participants may have had exposure to these concepts already, we wanted to ensure all participants had the required background, including how to implement these constructs in SNAP!

The video had three parts. The first part described the topic of variables, including how to create a new variable, as well as how the ‘set’ and ‘change’ blocks functioned: set blocks set the value of a variable and change blocks change that value by a predetermined increment. The video also described how to create a variable that corresponded to a list of numbers. The second part of the video focused on looping constructs, by showing participants how to use a ‘repeat’ block to create a loop that would fill the list from the first part with random numbers. The video also showed how to output information to the screen through the ‘say’ block, which prints out whatever value is inserted into it. Finally, the third part introduced conditional statements to show participants how to check whether one score was smaller than another. The third part implemented the program to print out the smallest score in a list of numbers (see Figure 3.3). Thus, by the end of the video, all of the background needed for the sorting task had been covered.

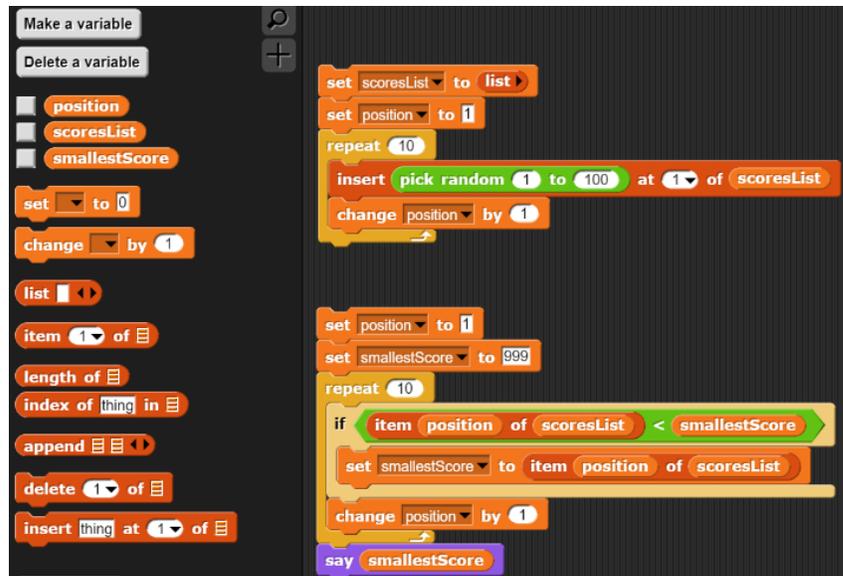


Figure 3.3. Tutorial Program

### 3.2.3.2 Sorting lesson

We also created a video containing a lesson about how to sort a list of numbers that was 23 minutes and 17 seconds in length. The video was narrated by the thesis author and had two parts, conceptual foundations behind sorting and implementation in SNAP! For part one, we first created a PowerPoint slide deck that was based on a lecture from CGSC1005 (the first-year programming class) but refined for the present study. The first slide introduced the sorting algorithm as a series of three steps:

1. find the position of the smallest element in the list
2. insert the element at that position into the end of a new list
3. delete the element at that position from the original list

The next 5 slides showed a worked example of how the algorithm could be solved on ‘paper’ (see Figure 3.4). The last slide detailed the four big ideas listed above (see section 3.2.2) as well as some contrasting cases where those ideas were not adhered to and the solutions became incorrect. For example, we showed participants what would happen if the element that was inserted into the sorted list was not deleted from the

original list. This would cause the program to continuously find and insert that element and no other. In the second part, the video described how to implement the sorting algorithm in SNAP! This involved creating a nested loop, where the outer loop was responsible for resetting position counters and inserting/deleting elements as needed, while the inner loop was responsible for finding the position of the smallest element. Inside the inner loop there was an if statement that checked if the element at a given position was smaller than all other elements in the list. Once the position of the smallest element was identified, the program changed the value of the variable that kept track of smallest position to the position of the element found above, which was then inserted into the new (sorted) list. Both parts of the video were scripted beforehand so that the recording process was streamlined and all concepts were covered.

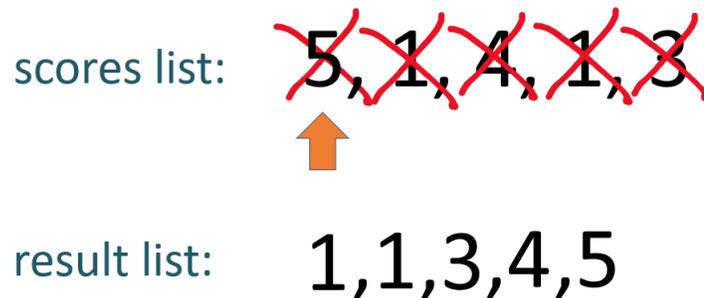


Figure 3.4. Worked example of sorting. In the instructional video, animations were added to show each element's removal.

### 3.2.4 Cheat sheets

To help participants with details related to SNAP! we created a cheat sheet that included the definition for every SNAP! block relevant for solving the sorting problem as well as some more general information about SNAP!'s interface (see Appendix B.2). To help guide the participants, a second cheat sheet was made that had a picture of the

SNAP! interface with labels so that the participants could use the labels to guide their collaboration (see Appendix B.1).

### 3.2.5 Pretest and posttest

We designed a pretest and posttest to measure both procedural and conceptual knowledge. The pretest consisted of two questions: (1) a code tracing question that gave participants a code snippet and asked what the output would be and (2) a question asking participants to provide an algorithm (in English/code/pseudocode) to insert all numbers in a list between 20-30 into a new list<sup>2</sup>. Both of these questions were designed to test participants' procedural knowledge; the first tested their ability to read code and the second tested their ability to write code.

The posttest consisted of three sections: (1) the two questions from the pretest (referred to as '*procedural*'), as well as one question asking participants to replicate the sorting algorithm from the programming lesson (to gauge retention from the experiment) (referred to as '*recall*'), and seven questions about the concepts underlying the sorting algorithm (referred to as '*conceptual*')<sup>3</sup>. The seven questions were used to measure participants' conceptual understanding related to sorting. For example, one of the questions asked: "Why does the program need to set `minPosition` and `position` to 1 each time the outer loop cycles around (see red arrow showing outer loop) – why not just set them to 1 once outside the outer loop?" The purpose of this question was to test whether participants had grasped the concepts underlying looping through a list, one of which is

---

<sup>2</sup> The first two participant pairs had an additional two questions. These were cut to reduce time taken to complete the study.

<sup>3</sup> The first two participant pairs had an additional two questions on their posttest. These were cut to reduce time taken to complete the study.

that the program should start at the beginning of the list each time so that some elements are not skipped over (see Appendices A.1 and A.2 for the pretest and posttest respectively).

The pretest and posttest were not identical for two reasons: participants' (lack of) prior knowledge and timing constraints. In terms of timing, the pretest being as long as the posttest would have made the experiment take longer than two hours, which we wanted to avoid so as not to fatigue participants. In terms of prior knowledge, we recruited participants without in-depth programming knowledge. Given that sorting is a more advanced topic, asking them about sorting on the pretest would not be productive.

### **3.3 Materials: Questionnaires**

The study involved a variety of questionnaires that can be found in Appendix C.

#### **3.3.1 Demographic questionnaire**

The demographic questionnaire consisted of four questions (measuring age, gender, native language, and area of study).

#### **3.3.2 The academic emotions questionnaire (AEQ)**

The AEQ (Pekrun et al., 2011) measures the emotions that arise in three separate academia-related contexts: learning, studying, and test-taking. The questionnaire contains subscales measuring enjoyment, hope, pride, relief, anger, anxiety, hopelessness, shame, and boredom. Each subscale of the questionnaire relates to one emotion/context pairing (e.g. learning-related enjoyment). The subscales are comprised of statements pertaining to that pairing and are rated on a five-point Likert scale (1= completely disagree-5= completely agree). Because the questionnaire is long (232 questions total) and because our study focused on a learning context and certain target emotions, we only used the

following subscales: Learning related enjoyment, hope, pride, anger/frustration, anxiety, shame, hopelessness and boredom (75 questions total).

### **3.3.3 The self-efficacy questionnaire**

The self-efficacy questionnaire (Sherer et al., 1982) measures a participant's perception of their ability to master everyday tasks and succeed at them. This questionnaire is comprised of two subscales: the general self-efficacy scale (17 items) and the social self-efficacy scale (6 items). We did not use the social self-efficacy scale seeing as it was unrelated to the context of the experiment. The original questionnaire used a 14-point Likert scale that we adapted to a five-point scale in order for it to be similar to the other scales.

### **3.3.4 The curiosity questionnaire**

The curiosity questionnaire (Litman and Spielberger, 2003) consists of 56 items split into two categories: 40 measuring epistemic curiosity and 16 measuring perceptual curiosity. Epistemic curiosity is defined by Berlyne as "drive to know" and is evoked by knowledge gaps and conceptual puzzles (Litman & Spielberger, 2003). All 40 items were rated on a four-point Likert scale from 1=almost never - 4= almost always. Perceptual curiosity pertains to the physical world around oneself and one's perception of it. As such, those 16 items were not relevant and so were not used in the present experiment.

### **3.3.5 The intolerance for uncertainty scale**

The intolerance for uncertainty scale (Freeston et al. 1994) measures participants' acceptance (or lack thereof) of uncertainty in everyday situations. The scale is comprised of 27 items that measure if feelings of uncertainty are acceptable and/or lead to frustration, stress, and the inability to act (Buhr & Dugas, 2002). All items are rated on a

five-point Likert scale ranging from 1= 'not at all characteristic of me' to 5= 'extremely characteristic of me'.

### **3.3.6 The experience sampling questionnaire**

The experience sampling questionnaire consisted of five questions asking participants how they felt based on a five-point Likert scale (1=not at all –

5=extremely). The emotions measured included anxiety (negative valence, high arousal), boredom (negative valence, low arousal), enjoyment (positive valence, moderate arousal), frustration (negative valence, arousal depends on the participant), and confusion (positive or negative valence, arousal depends on the participant).

### **3.4 Design and Procedure**

The study was a single-group design with all participants receiving the productive failure intervention. While we originally planned a two-condition study comparing productive failure and direct instruction, we decided to focus on the productive failure condition given that the pandemic hit at the very initial stages of recruitment. Given this development, we moved our study online and since this was a novel development, there was uncertainty about how many participants we could recruit. Thus, we shifted our focus from an experimental to a single-condition study. In retrospect, while this was not our original plan, to the best of our knowledge this is the first study to collect data on student emotions in a productive failure study and given this preliminary stage, an exploratory approach is suitable.

With the exception of the first three sessions (details on these are below), each study session was conducted with a single pair of participants over a two-hour period using Zoom (i.e., all sessions were virtual and included the experimenter and two

participants)<sup>4</sup>. The experiment began with the experimenter introducing himself and asking participants to provide informed consent<sup>5</sup> by filling out the consent form sent to them over Zoom chat. Once consent was provided, the two participants were reminded that the experiment would consist of individually answering some questionnaires and collaborating on a problem-solving task (this information was also in the consent form). Each participant was then given an online link to the demographic, AEQ, self-efficacy, curiosity, and intolerance for uncertainty questionnaire. After individually completing these questionnaires, each participant was given the link to the online pretest (with a 10-minute time limit) and they individually completed this test (they were told to write ‘IDK’ if they did not know the answer to a question). Following the pretest, the experimenter shared their screen with both participants and showed the 20-minute tutorial video to introduce participants to SNAP! and the basic programming features they would need for the upcoming productive failure problem. Participants could watch the video but could not pause/rewind/forward it (as it was shown on the experimenter’s screen).

Once the video had concluded, the productive failure phase began during which participants worked on the sorting task collaboratively. Participants were given the SNAP! template via a Google Drive download link that they opened on their computer. The template contained a snippet of code that generated a random list of numbers - participants were asked to write a SNAP! program to sort that list from greatest to least or

---

<sup>4</sup> The first three pairs were in the same location; subsequent sessions were conducted over Zoom. Unfortunately, we do not have data on whether participants in a given pair were co-located, but given the social distancing rules imposed by the pandemic, we anticipate that most were not.

<sup>5</sup> Participants were asked if they were current students of Dr. Muldner. If they were, they were sent a modified consent form that informed them that their grades in those classes would not be affected by their participation. If they were not, they were given the standard Carleton consent form (See Appendix E). In total, only two participants required this modified consent form.

least to greatest (they were given 20 minutes to work on this problem). To help with SNAP! logistics, participants were given the cheat sheet containing all the SNAP blocks and their definitions, as well as the second cheat sheet with a snapshot of the interface.

Because this was a collaborative task, the participants were asked to work together and talk to each other. The instructions given to participants were as follows:

*“You will have 20 minutes to **work together** to solve the problem by sorting the list generated by the code snippet from either least to greatest or greatest to least. You may not use any notes other than the provided cheat sheets. The experimenter will also answer no questions that are related to the problem. If SNAP! for some reason stops working or something abnormal happens with the interface, those questions will be answered. Because we will be recording this task, please turn off your webcams and make sure to speak aloud when solving the problem. I will let you know when there are 10 minutes left. Good luck!”*

Due to the online nature of the study, one participant was asked to share their screen with the template on it - the other participant could see this screen but could not directly work in the interface and so needed to tell their partner what SNAP! actions they wanted them to perform. While we originally planned for both participants to have physical access to this interface, this was not logistically possible. However, this type of situation where one participant is performing the physical actions and the other contributing both by presenting ideas and guiding some actions is one that happens naturally in various educational settings like labs where one student is in charge of the keyboard.

The recording of the problem-solving task was done through Zoom and because the sessions were recorded and later analyzed, for privacy reasons, participants turned off their webcams during this part. Once 20 minutes had passed, participants were told to stop and individually answer the experience sampling questionnaire for the first time. Next, the experimenter shared their screen again to show participants the lesson video that detailed the solution (both high-level and implementational) to the problem. Once they had watched the lesson video, they answered a second experience sampling questionnaire. Finally, participants were given the posttest, which they had 20 minutes to individually complete. Once done the posttest, they were told to individually answer the experience sampling questionnaire for a third and final time. The participants were then debriefed.

The above procedure was followed for all participants but with a few variations for the first three pairs, who completed the study in the lab (these three pairs were recruited before the pandemic restrictions occurred). The procedure differences were as follows: (1) the pretest and posttest were on paper as opposed to online, (2) the participants watched both videos and completed the task on one computer sitting next to each other, which allowed for shared control of the mouse/interface. The original procedure was piloted with one other participant and any necessary adjustments were made; once the shift to online occurred, the key parts were piloted with two individuals.

### **3.5 Qualitative Coding**

We analyzed emotional expression as well as expressions of failure and ideas during the productive failure phase using a qualitative coding approach. As mentioned above, all sessions were recorded using Zoom's built-in recording feature. These

recordings were subsequently transcribed by the thesis author and coded. To prepare for coding, an initial coding scheme was constructed.

The coding scheme included the following constructs: (1) frustration; (2) confusion; (3) uncertainty; (4) failure and (5) ideas. Ideas and failure were chosen because they would be indicators of progress (or lack thereof). The three emotions were chosen because, in a paradigm that involves failure, we speculated that participants would experience more negative emotions. Specifically, we speculated that levels of confusion, frustration, and uncertainty would be higher in this paradigm as compared to a standard learning paradigm. Therefore, we focused more on those negative emotions as opposed to positive emotions (e.g. enjoyment or pride) for the coding scheme.

An initial coding scheme was developed for these constructs, including details on verbal expressions indicative of each construct, and was refined by coding the first three transcripts. For segmentation purposes, we defined an utterance as a phrase spoken by a participant. We defined a speaking turn as all the utterances made by a single participant in a given turn. This segmentation allowed us to code the constructs at the utterance level, which was a more fine-grained approach than at the speaker level. The thesis author and supervisor independently coded the first transcript and compared their results in order to improve the coding scheme. The coding scheme was refined through discussion with the thesis supervisor and this process was repeated for the next two transcripts – the final coding scheme is shown below. The coding scheme was then used by the thesis author to code the other 20 transcripts. After the initial coding was done, the thesis author re-read the transcripts to ensure accuracy and completeness. The quantitative data for each code (i.e. the number of expressions of each code per participant) were tallied and entered into

SPSS on a per participant basis i.e. each participants' coded instances were counted separately.

### 3.5.1 Coding Scheme: Details

The following is a detailed coding scheme that describes how each construct is defined and gives examples of each construct. All coding was done at the utterance level so that we could capture more than one code in a single speaking turn (e.g. if participants expressed more than one idea per speaking turn). The constructs and their descriptions are below.

**Frustration:** a participant expressed irritation and/or they were having a hard time getting a particular part of the program to work. This is based on the standard definition, namely that frustration is "...a key negative emotion that roots in disappointment and can be defined as irritable distress in response to limitation, exclusion, and failure" (Jeronimus & Laucelle, 2017). Some examples of frustration from the data are: "*I don't understand why this isn't working!*", "*What the heck is going on?*", "*Why are we so stuck?*".

**Confusion:** a participant expressed that they were unsure about how to proceed or how to move forward towards a solution. This is based on the standard definition, namely that "cognitive disequilibrium and confusion are triggered when individuals encounter incongruence in the form of impasses, anomalies, contradictions, disruptions of goals, extreme novelty that cannot be comprehended, and interruptions of organized sequences of actions... Confusion is thought to be the affective signature of cognitive disequilibrium" (D'mello & Graesser, 2014). Participants either did not have an idea/plan on how to move forward or they were explicitly expressing confusion. Some examples of

confusion are utterances such as: “*I am so confused right now*” or “*I have no idea where to go from here*”.

**Uncertainty:** a participant expressed ideas that they were not sure were correct; they had an idea but were uncertain about its validity. This was based on the definition of uncertainty, namely that it is “...a dynamic state in which there is a perception of being unable to assign probabilities for outcomes that prompts a discomfoting, uneasy sensation” (Penrod, 2001). Confusion and uncertainty are distinct in the present scheme because being in a state of confusion implies that one does not have any ideas about how to proceed whereas, when one is uncertain, they have ideas about how to proceed and do not know which one, if any, are correct. Some examples of this code are: “*I’m not sure this will work but...*”, “*Could we maybe try this?*”, “*I don’t know whether we’re able to do this or that...*”. Additional notes:

- Questions were coded as uncertainty if they were not directly seeking validation or confirmation (their aim is to explain something or express uncertainty). Clarifications are not indicative of uncertainty.
- For example, a question such as “*Why would you not want to do X?*” is expressing uncertainty as to whether X should be done or not done. This type of question was coded.
- On the other hand, a clarifying question such as “*Sorry did you say put two loops in there?*” was not coded because the participant was asking for clarification on something their partner had said and was not expressing uncertainty concerning an aspect of the problem.

**Failure:** a participant recognized that they did not reach a solution for a specific part of the problem or for the whole problem, either by expressing that they were stuck or that they had made a mistake. Some examples of failure are utterances such as: “*that didn't work*” or “*whoops that's not right*”.

**Idea:** a participant expressed an idea on how to solve the problem (even if it was incorrect). Some examples of ideas include: “*I guess we could set position to one right// because it's variable*”, “*We can create another variable that's just basically following position*”, “*we could just make a temp variable*”. Notes:

- Given that ideas were coded at the utterance level and not at the speaking turn, a participant could express more than one idea in an utterance (for example: “*Set it to item position of score list //...// Then set item position of score list to item position plus of score list*”). The slashes here are used to denote the separation between ideas.
- Verbalizations describing what was being done or a known fact were not coded for. For example: ‘why don't we try X’ was coded for whereas statements such as ‘we know that Y works’ or ‘this is Z numbers long’ were not.
- All ideas were counted unless they were repeated in quick succession. By quick succession, we mean that the same idea (in terms of content) was repeated within two lines of being uttered for the first time. If the idea was repeated with novel content added in, that new idea was coded.

## **Chapter 4: Results**

The analysis was guided by the following research questions: (1) What is the frequency of occurrence of frustration, confusion, and uncertainty during productive failure? (2) How do emotions change based on instructional activity? (3) What is the relationship between emotions and learning? To answer the first question, we use a qualitative approach to describe how emotions and ideas were expressed during the productive failure phase. To address the latter two questions, we report both descriptive and inferential statistics. For analyses involving statistical tests, the associated assumptions were checked, and any violations are reported.

### **4.1 Emotions Expressed During Productive Failure**

We begin with a description of the target constructs that were verbally expressed by the participants during the productive failure phase. The recordings from the productive failure phases were transcribed and coded as described in section 3.5. We were interested in how much each target emotion was expressed (recall we coded for confusion, frustration, and uncertainty), as well as the expression of ideas and failure. Table 4.1 shows the results for these constructs, while Figure 4.1 shows the data for each construct in a graphical form. These figures include data from 46 participants (one pair was dropped due to recording issues). This pair did however, complete the rest of the study and was included in the rest of the analyses.

In general, emotion expression was not frequent in the verbal protocols. As shown in Figure 4.1 on average, each participant expressed frustration less than once per entire session – confusion and uncertainty were more common, being expressed on average

around three times per participant; a similar result was found for the expression of failure. In contrast, participants did express many ideas during productive failure.

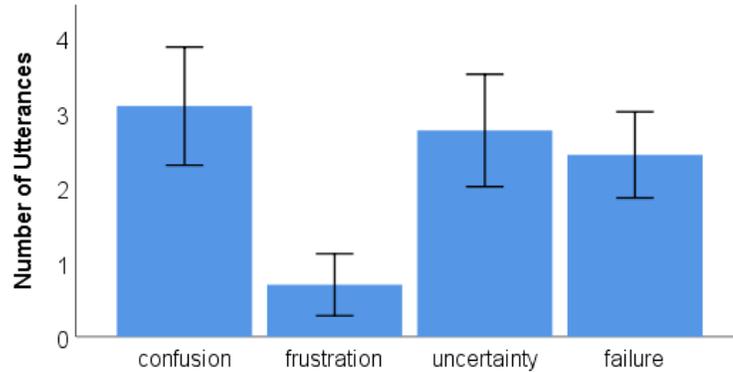


Figure 4.1. Mean number of utterances for confusion, frustration, uncertainty, and failure, per participant. Ideas were not included as they occurred more and affected the scale of the graph (error bars represent 95% confidence intervals)

Table 4.1: Mean and standard deviation of expressions related to emotions, failure, and ideas

Variable	Mean	SD
confusion	3.09	2.67
frustration	.70	1.40
uncertainty	2.76	2.53
failure	2.43	1.94
ideas	27.70	15.82

*N* = 46 (one pair was dropped due to recording issues).

#### 4.1.1 Examples of emotions expressed during productive failure

We now provide a qualitative description of examples of emotional expression, as well as failure and ideas, which occurred in the verbal protocols (transcripts). Below, the protocols are labeled by their ID (e.g. PF109); each protocol corresponds to two participants and participants are labeled with their protocol ID and participant label, either A or B (e.g. PF109A).

**Frustration:** Frustration was commonly linked to lack of progress and/or failure. To illustrate, PF117's participants expressed frustration at the lack of progress they were making. These participants had a little experience with Python (a commonly used programming language) but not with SNAP! and ran into difficulty when they tried to store a value in a temporary variable and subsequently delete the value in the list at that position. This solution relied on accessing positions in the list but because of incorrect logic, their program accessed a position that did not exist (position 11 in a list with 10 entries). This caused the program to behave erratically (erasing values and/or duplicating certain values). Aggravated by their mistakes, PF117B said "*Oh, I wish this was any other code language*" and PF117A replied with "*Give me Python. I want to go back to Python*".

As a second example of frustration, PF118's participants had been trying to solve the problem for 18 minutes and had made little progress. After several consecutive failed runs of the program, PF118A said "*Could you imagine if that actually worked? I'd cry. I think I'm probably gonna cry.*"

As a third example of frustration, PF112's participants had been working on their program for almost 17 minutes and had some of the correct components needed to solve the problem. However, they were trying to incorrectly extract a value from their list using the value itself as opposed to its position. When the participants ran their program for the second time (after having made an inconsequential adjustment to a variable), they found that it did not work again and PF112A exclaimed "*C'mon, work!!... Why won't you work?*".

**Confusion:** Confusion corresponded to phrases such as: “*I’m confused*”, “*What’s going on*” or “*Why didn’t that work*”. We will illustrate confusion using PF105’s protocol. When these participants first attempted to run their program, it inadvertently created a list containing the same value, repeated 10 times. This was because their loop inserted the first value smaller than 999 into the list 10 times and there was no code that would sort the values. When they saw that their list had changed from an unsorted list of 10 different values to a list that contained the value 69 10 times, PF105A said “*Why didn’t that work?*”. PF105B expressed similar confusion by saying “*Why did they all say 69?*” They were confused because this was not the intended result. This same error and subsequent confusion occurred in PF125’s transcript.

Another example of confusion occurred in PF109’s transcript when participants were testing their program. The first attempt was incorrect because the program only ran 9 times due to a flaw in their loop condition. Moreover, their two ‘if’ blocks were incorrect. Both blocks were designed to insert a value into the new sorted list if it was smaller than a pre-stored item. This did not go according to plan because their position variables, used to index the values in the list, were incorrect. When they first ran the program, it did not work and so, one participant suggested nesting one of the ‘if’ statements inside the other. When they tried running the program again, the resulting list had one blank and 9 unsorted entries as opposed to 10. When they saw this result, PF109A said “*What? No, now, it's nine. And there's an empty one.*” This is an example both of failure and of confusion. The participants failed to create a program that produced the desired result, and they were confused about the blank space in the resulting list.

**Uncertainty:** Uncertainty was often characterized by phrases such as “*I don’t know*” or “*I’m not sure*”. Recall that uncertainty and confusion are distinct in that uncertainty implies that participants had an idea about how to proceed but are not sure about whether it is correct, while confusion implies that participants do not know how to proceed. We will illustrate uncertainty with PF119’s transcript. When trying to build their ‘if’ block to find the smallest value in the list, they created a variable called position. PF119B thought that position could be used to keep track of the spot in the list but expressed doubt about this proposal: “...*I’m pretty sure we’re gonna be using position for something else but I’m not sure*”. Here participants are expressing uncertainty about what they will use position for. Later on, PF119B expresses uncertainty again but this time, they are unsure about why they added something to their program “...*Set smallest to increasing. I don’t remember what this is for...*”.

A second example of uncertainty is in PF101’s transcript. Early on, before they started to build their program, they were looking over the SNAP! blocks available to them and PF101A suggested that they will need a change block. PF101B agreed but was uncertain about how the change block should be used. They said “*But the smallest number will stay the same. How’s it gonna get like the next smallest?*”. Here the participant is expressing uncertainty about how a block that changes variable values would be able to keep track of a number that constantly changes. They are aware that the block will have to be used, they are just unsure of its use.

**Failure:** Failure corresponded to utterances like “*that’s not right*”, “*that doesn’t work*” or “*oh no*”. One example of failure occurred in PF118’s protocol. When trying to build their program to find the smallest element, they originally had their program check

whether ‘*item 1 of scores list*’ was less than ‘*item 2 of scores list*’. This is incorrect for two reasons: (1) one of the main ideas behind the sorting algorithm is that we need to deal with positions and not values and (2) this method will only check two positions in the list but not the rest. PF118B said “*You can't make 10 of these... Item one isn't the same as position*”, recognizing that there is an important distinction between position and value of elements; PF118A replied “*All right. That's an issue*”.

As another example of failure, when PF110’s participants were trying to sort their list, their approach was to add several insert blocks after the loop, each of which would put a random number at position 1 of the new list. Participants realized this was incorrect when they ran their program and saw that the new list contained three random numbers as opposed to the list being sorted. PF110A realized this strategy was not the answer and so concluded that “*This is not the way...*”.

We provide a third example of failure using PF122’s protocol. After several minutes of trying to find a way to solve the problem, the participants created a loop and inserted an ‘if’ block inside to try and find the smallest item in the list. These steps would have led to at least a partially correct solution, but the participants thought that the code that created the initial list and populated it with random numbers had to be put into the loop. This led their program to just make several unsorted lists, one after another. PF122A commented “*No it doesn't do much...Nope, didn't do anything*” and then followed her realization by exclaiming “*Yeah I'm stuck*”.

**Ideas:** Expression of ideas often involved ‘trying out’ a method or fix to the program. One example of this came from PF108’s protocol. When trying to figure out how to create their ‘if’ statement for the sorting algorithm, the participants dragged in a

'item \_\_\_ of \_\_\_' block to tell the program which element to check. PF108B said "*Try to drag the positions into that slot*" expressing an idea (i.e., that in order to find the smallest element, they needed to compare the value of 'smallest number' to an item at a certain position of the list). Other examples of ideas that were common were: "*we'll have to do it 10 times*" and "*Let's just create a new list*", with variations of both being present in many transcripts (15 out of 23).

There were various instances where participants had ideas that flowed into one another. One such instance was in PF111's transcript where PF111A realized that a series of steps had to be taken in order to start the sorting process. They proposed that "*...then we can have it sorted//, but we have to do that first step right, the one that he showed us of identifying and displaying what the lowest or greatest number is. So, we need to make a variable called smallest number, right?*" (the '// shows the segmentation of the different ideas). Here, the participant realized that they needed to sort from the original list. To do so, they would need to find the smallest number and in order to store that number, they would have to create a variable.

Some of ideas, such as the one from PF111A above, were incorrect. While these were still coded as ideas because we did not differentiate based correctness, the incorrect ideas were not directly helpful for participants' solutions. For example, in PF123's transcript, participants had started building their program and had a loop with a nested if block to repeatedly check for the position of the smallest element. Unfortunately, they confused position and value. PF123B was certain that they needed to "*...have to have smallest score and we have to set it...to a big number*". This is not correct as that would

find the smallest value rather than the position of that value. However, incorrect ideas may have been the catalyst for subsequent correct ideas.

## **4.2 State and Trait Emotions**

Above, we described the occurrence of emotions, failure, and ideas during the productive failure phase. To examine participant emotions over the course of the entire study, we used data from two other sources: (1) the emotion questionnaire presented at the beginning of the experiment and (2) the experience sampling questionnaire presented at three timepoints. Table 4.2 shows the results for the AEQ, which measured trait academic emotions and Figure 4.2 displays them graphically. Table 4.3 shows the results for the related traits (self-efficacy, curiosity, and intolerance for uncertainty). Scores for all three traits fell above the middle value in the range of possible scores for each construct.

The experience sampling questionnaire collected self-reports on emotions at three time points (right after the productive failure phase, after the lesson, and after the posttest). The results for the state emotions are shown in Table 4.4 while Figure 4.3 graphically shows the trends in emotional change over time. It is important to note that while the emotions are graphed together in Figure 4.3, the levels of one were not compared to another as that comparison is irrelevant to what we wanted to examine.

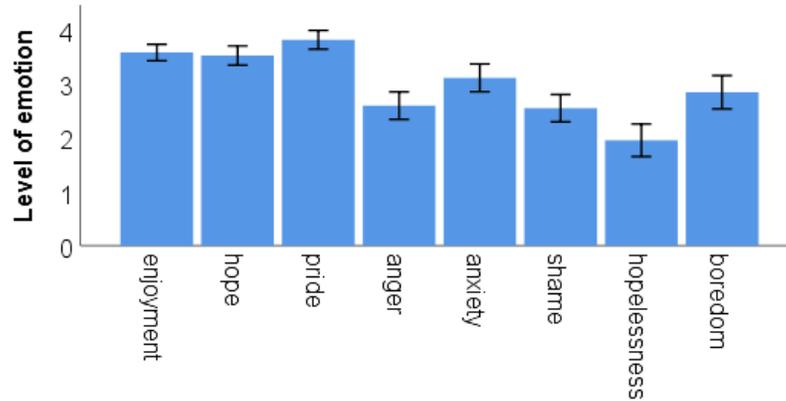


Figure 4.2. AEQ trait academic emotions (error bars represent 95% confidence intervals)

Table 4.2: Mean and standard deviation for the AEQ trait academic emotions (/5)

Variable	Mean	SD
enjoyment	3.57	.52
hope	3.51	.61
pride	3.80	.59
anger	2.59	.88
anxiety	3.10	.88
shame	2.54	.87
hopelessness	1.95	1.03
boredom	2.83	1.06

$N = 48$

Table 4.3: Mean and standard deviation for the other trait academic emotions

Variable	Mean	SD
self-efficacy <sup>a</sup> (/85)	59.81	10.60
curiosity <sup>b</sup> (/160)	119.96	16.38
intolerance for uncertainty <sup>a</sup> (/135)	74.13	18.97

<sup>a</sup>  $N = 48$ . <sup>b</sup>  $N = 47$  (one participant did not fill out the questionnaire)

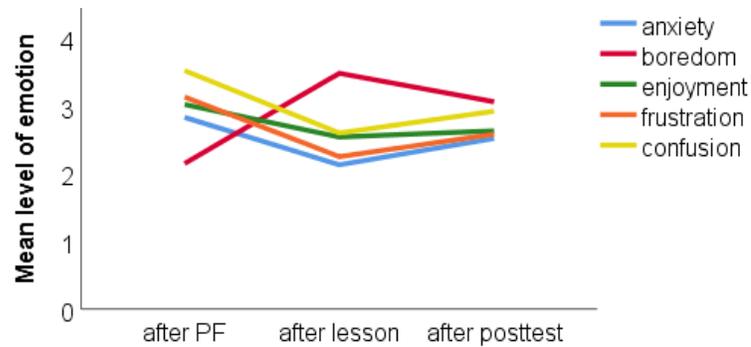


Figure 4.3. Change in emotion (possible range 1 – 5) over the three target timepoints for each of the five target emotions (error bars not shown because they made the figure unreadable – see Table 4.4 for the variance information for these variables)

Table 4.4: Mean and standard deviation for the experience sampling data (shown as mean followed by standard deviation in brackets)

timepoint	anxiety	boredom	enjoyment	frustration	confusion
after PF	2.83 (1.31)	2.15 (1.22)	3.02 (1.02)	3.12 (1.30)	3.52 (1.43)
after lesson	2.13 (1.23)	3.48 (1.15)	2.54 (1.01)	2.25 (1.21)	2.60 (1.33)
after posttest	2.52 (1.23)	3.06 (1.14)	2.63 (.94)	2.58 (1.38)	2.92 (1.32)

$N = 48$

In order to investigate patterns of emotion over time, we used five separate repeated measures ANOVAs with emotion as the dependent variable and timepoint as the repeated factor to examine effect of timepoint on each emotion (anxiety, boredom, enjoyment, frustration, and confusion). Two outliers were present for boredom and one was present for enjoyment, but they were legitimate data points and when excluded, the results were not affected. Thus, the outliers were not removed. When testing for normality, the Shapiro-Wilks test showed violations of the assumption for all emotions (see Appendix D for skew and kurtosis values). Repeated measures ANOVAs, however, are robust to violations of normality (Lix et al., 1996) and so we still conducted the

analysis. Sphericity corrections were made using the Huynh-Feldt correction. We followed significant effects with trend analysis because this sheds light on patterns over time for the target dependent variable. If several trends were significant or marginally significant, we only reported the highest order polynomial. The three timepoints at which emotion was measured are shown in Figure 4.3; recall that timepoint one was after the productive failure phase, timepoint two was after the lesson, and timepoint three was after the posttest.

There was a significant main effect of timepoint on boredom,  $F(2,94) = 29.16, p < .01, \eta_p^2 = .38$ , indicating that participants' level of boredom was affected by the activity that directly preceded the emotion self-report. The data followed a quadratic trend,  $p < .01$ . Boredom was lowest right after the productive failure phase, highest after the lesson, and then decreased slightly after the posttest. This shows that students were more interested in the productive failure phase compared to the lesson.

There was a significant main effect of timepoint on anxiety,  $F(2,94) = 11.09, p < .01, \eta_p^2 = .19$ , indicating that participants' level of anxiety was affected by the activity that directly preceded the emotion self-report. As with boredom, the data followed a quadratic trend,  $p < .01$ , but the pattern was different. Anxiety started off high after productive failure but dropped considerably after the lesson, increasing again after the posttest. The other emotions (enjoyment, frustration, and confusion) followed the same pattern as anxiety with a decrease between the end of the productive failure phase and the lesson and then increasing again after the posttest (enjoyment:  $F(1.72, 80.6) = 5.79, p < .01, \eta_p^2 = .11$ , marginal quadratic trend,  $p = .053$ ; frustration:  $F(2,94) = 15.73, p < .01, \eta_p^2$

= .25, quadratic trend,  $p < .01$ ; confusion:  $F(1.75, 82.37) = 14.89$ ,  $p < .01$ ,  $\eta_p^2 = .24$ , quadratic trend,  $p < .01$ ).

In summary, all of the emotions changed significantly over time. This shows that all three instructional phases (productive failure, lesson, and posttest) had distinct effects on participants' emotional states; high-arousal emotions (anxiety, confusion, frustration, and enjoyment) all decreased after the lesson and increased after the posttest while the low-arousal emotion (boredom) increased after the lesson and decreased after the posttest.

### **4.3 Learning and Emotions**

Learning is a broad construct. Here, we operationalize it as follows: procedural learning is the change from pretest to posttest and conceptual learning is measured using the scores on the conceptual posttest (for conceptual learning, we did not have pre-test scores - we acknowledge that the present measure does not capture the change in knowledge, but since our participants were novices and not likely to have sorting expertise, the post-test scores approximate this construct). We begin with the results related to learning and then present results on the relation between learning and emotion. To investigate whether participants learned from the productive failure intervention, we graded the pre and posttests using a rubric (see Appendix A3) and computed the differences between them. As described in Section 3.2.5, the pretest consisted of two procedural questions, while the posttest consisted of three parts, namely procedural questions, the sorting algorithm recall, and conceptual questions, referred to as *procedural*, *recall*, and *conceptual*. The total score was based on the sum of the scores for

questions corresponding to these three parts and is referred to as *postTotal*. Table 4.5 and Figure 4.4 show the scores for each test component as well as the total score.

Overall, the posttest scores were low, with many participants scoring below 50% (procedural: 18/48 scored below 50%; recall: 33/48 scored below 50%; conceptual: 22/48 scored below 50%). In particular, overall conceptual learning was low. However, while performance was low, students did learn. A paired samples t-test showed a significant increase in scores from pretest,  $M = .65$ ,  $SD = 1.00$ , to procedural,  $M = 1.73$ ,  $SD = 1.07$ ,  $t(47) = 8.44$ ,  $p < .01$ ; the effect size,  $d = 1.21$ , was large. In this analysis, recall and conceptual test sections were not included because they measured conceptual knowledge, something that was not part of the pretest for reasons described in Section 3.2.5. In contrast, the pretest and procedural test section measured procedural knowledge.

Table 4.5: Mean and standard deviation for the pretest and posttest scores (shown as raw scores followed by percentages in brackets)

Variable	Mean	SD
pretest (/3)	0.65 (22%)	1.00 (33.36)
procedural (/3)	1.73 (58%)	1.07 (35.88)
recall (/5)	1.54 (31%)	1.40 (28.05)
conceptual (/10)	4.66 (47%)	2.56 (25.62)
postTotal (/18)	7.91 (44%)	3.97 (22.08)

$N = 48$

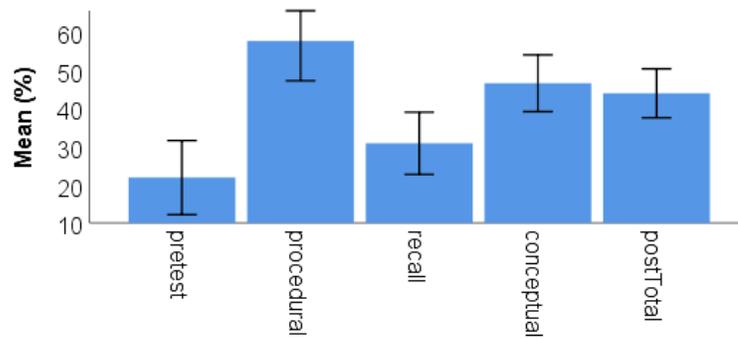


Figure 4.3. Scores for pretest, as well as the procedural section, recall section, conceptual section, and postTotal (sum of scores) (error bars represent 95% confidence intervals)

Another construct potentially related to learning is the number of ideas participants generated during productive failure. Controlling for pretest, we did not find evidence that the number of ideas produced was related with postTotal,  $r = .21, p = .174$ . This shows that the number of ideas expressed by participants during the productive failure phase did not change their posttest performance.

#### 4.3.1 Relationships between emotions and learning

In this section, we report on the relationship between emotions and learning. We focus on the trait emotions measured before the productive failure stage and the state emotions measured by experience sampling as the emotions identified in the verbal protocols were too infrequently reported to warrant quantitative analysis. To analyze the relation between trait emotions and learning outcomes, we ran exploratory correlations between the eight AEQ emotions (enjoyment, hope, pride, anger/frustration, anxiety, shame, hopelessness, and boredom) as well as the three related traits (self-efficacy, curiosity, and intolerance for uncertainty) and total posttest scores (labelled as postTotal in Figure 4.4), controlling for pretest (see Appendix D.2 for results).

We found no significant relationship between trait emotions and posttest scores. This was unexpected because previous studies have shown that these emotions do relate to learning outcomes (see section 2.3). We return to this result in the discussion.

We now turn to the results on the state emotions. Given that we had three measures for each emotion, one per time point (referred to as: anxiety1, anxiety2, anxiety3 for anxiety – the same convention was used for the other emotions), we used a linear regression analysis. Specifically, we ran a series of regressions, one for each emotion, with the total posttest score as the outcome variable (i.e., postTotal) and as predictors each of the three emotion reports corresponding to the three timepoints they were measured at, as well as the pretest score. All three emotion/timepoints were used as predictors because we did not want to lose information by averaging the three measures. Here, regression was used rather than a series of partial correlations to avoid running multiple analyses and to examine the combined effects of multiple predictors at once.

One concern is that the predictors cannot be highly related, referred to as multicollinearity in the context of regression. To check for multicollinearity, we checked if the relation (Pearson  $r$ ) between any two emotion/timepoint measures was greater than .8, which would indicate multicollinearity (Vatcheva et al., 2016). This was not the case for any measure. We also checked that VIF was in the acceptable range ( $< 5$ ), which it was for all measures. Thus, we proceeded with the analysis, using the enter method.

The same pattern was found for four out the five emotions (anxiety, boredom, frustration, and enjoyment): the overall model was significant for each analysis, but the only significant predictor of posttest score was the pretest. For anxiety, the overall model was significant,  $F(4, 43) = 6.31, p < .01, R^2 = .37, R^2_{Adjusted} = .31$ . Anxiety was not a

significant predictor (anxiety1,  $\beta = .15$ ,  $p = .40$ , anxiety2,  $\beta = .06$ ,  $p = .72$ , and anxiety3,  $\beta = -.21$ ,  $p = .29$ ), while pretest did significantly predict postTotal scores,  $\beta = .59$ ,  $p < .01$ . For boredom, the overall model was significant,  $F(4, 43) = 7.53$ ,  $p < .01$ ,  $R^2 = .41$ ,  $R^2_{Adjusted} = .36$ . Boredom was not a significant predictor (boredom1,  $\beta = .14$ ,  $p = .30$ , boredom2,  $\beta = .07$ ,  $p = .65$ , and boredom3,  $\beta = -.30$ ,  $p = .058$ ), while pretest did significantly predict postTotal scores,  $\beta = .62$ ,  $p < .01$ . For enjoyment, the overall model was significant,  $F(4, 43) = 6.77$ ,  $p < .01$ ,  $R^2 = .39$ ,  $R^2_{Adjusted} = .33$ . Enjoyment was not a significant predictor (enjoyment1,  $\beta = -.03$ ,  $p = .84$ , enjoyment2,  $\beta = .09$ ,  $p = .57$ , and enjoyment3,  $\beta = .14$ ,  $p = .43$ ), while pretest did significantly predict postTotal scores,  $\beta = .58$ ,  $p < .01$ . For frustration, the overall model was significant,  $F(4, 43) = 7.51$ ,  $p < .01$ ,  $R^2 = .41$ ,  $R^2_{Adjusted} = .37$ . Frustration was not a significant predictor (frustration1,  $\beta = -.04$ ,  $p = .82$ , frustration2,  $\beta = .20$ ,  $p = .25$ , and frustration3,  $\beta = -.31$ ,  $p = .07$ ), while pretest did significantly predict postTotal scores,  $\beta = .55$ ,  $p < .01$ .

There was one exception to the above pattern: for confusion, as above, the overall model was significant,  $F(4, 43) = 7.81$ ,  $p < .01$ ,  $R^2 = .42$ ,  $R^2_{Adjusted} = .37$  and one of the confusion variables, namely confusion3 was a significant predictor, but the other two were not (confusion1,  $\beta = .03$ ,  $p = .84$ , confusion2,  $\beta = .18$ ,  $p = .34$ , and confusion3,  $\beta = -.41$ ,  $p = .03$ ). Pretest,  $\beta = .59$ ,  $p < .01$ , also significantly predicted postTotal scores.

In summary, the only significant predictors of posttest scores were pretest and the level of confusion participants reported directly after the posttest. These results will be discussed in the next chapter.

## **Chapter 5: Discussion**

Prior work has found that both emotions and productive failure impact student learning. The present work extended that research by exploring the presence and change in emotions in the productive failure paradigm. We found that productive failure, like other learning paradigms, does elicit academic emotions and that level of emotion is affected by instructional activity. Contrary to prior research, we did not find evidence of relationships between most of the emotions and learning.

### **5.1 Emotions and Related States in Productive Failure**

We recorded and analyzed participants' collaborative sessions and measured the presence of three different emotions and one related state during the productive failure phase: frustration, confusion, uncertainty, and failure. While we did find that all three of these emotions as well as expressions of failure were present in the protocols with some variability between participants (see Table 4.1), the overall occurrence of all the codes was low. Baker et al. (2010) also found similarly low levels of emotional expression during standard problem-solving activities. However, we anticipated that productive failure would elicit higher levels of the target emotions due to the nature of the paradigm. Specifically, productive failure requires that students work on open-ended problems that they are not expected to succeed in solving, which should produce high levels of confusion, uncertainty, and potentially frustration. One potential explanation for why this did not happen could be that the collaboration between the participants buffered the negative affect that usually accompanies failure. Pietarinen et al. (2018) found that there is a reciprocal, positive relationship between collaboration and positive affect. Thus, in the present experiment, participants' emotional states may have been positively valenced

because they were collaborating. In general, the collaborative aspect of the productive failure paradigm may diminish the negative emotions that accompany failure.

In addition to identifying the emotions present *during* productive failure, we also used experience sampling to collect information on emotion levels right after the productive failure phase, as well as subsequent instructional activities. This data showed moderate levels of enjoyment,  $M = 2.73$ ,  $SD = .79$ , and frustration,  $M = 2.65$ ,  $SD = 1.13$ , with levels of boredom and anxiety being slightly lower,  $M = 2.90$ ,  $SD = .93$ , and  $M = 2.49$ ,  $SD = 1.10$  respectively, and levels of confusion being slightly higher,  $M = 3.01$ ,  $SD = 1.18$ . Borracci et al. (2019) also measured boredom, frustration, and anxiety during standard problem solving (i.e., in contrast to productive failure, in the Borracci et al. study participants solved algebra problems with access to worked-out examples). They found that when students self-reported their emotions after solving 4, 8, and 12 problems, the reported levels of the three emotions were slightly below three out of five on a Likert scale ranging from (1= not at all to 5= extremely), which differed from our results. For example, in our study, the average level of anxiety after productive failure was 2.8 out of 5 whereas in Borracci et al., the level of anxiety after problem solving was 1.75 out of 5. Frustration after productive failure was higher in our study (3.13 out of 5) than in Borracci's work (1.7 out of 5). Finally, boredom in both studies was similar (2.5 out of 5 in Borracci's work and 2.15 out of 5 in this study). The higher levels of anxiety and frustration in the present study can be attributed to the fact that in Borracci's study, students had assistance for solving the problems (in the form of examples) whereas in our study, the participants did not have assistance, which may have led to higher levels of

anxiety and frustration (as well as confusion – but this was not measured in the prior study so we do not have a comparison).

Emotion was also measured at two other timepoints: right after the lesson and again right after the posttest. Four of the measured emotions, namely anxiety, confusion, enjoyment, and frustration, followed a similar quadratic trend. All four levels decreased after the lesson and increased slightly after the posttest. We speculate they decreased after the lesson because the lesson on how to solve the problem alleviated some of the confusion, frustration, and anxiety that resulted from not having solved the problem in the productive failure phase. Enjoyment's decrease after the lesson lead us to speculate that the productive failure phase was more enjoyable than the lesson. In contrast, boredom exhibited the opposite pattern of change over time: it was the lowest after the productive failure phase, then increased sharply after the lesson and finally decreased slightly after the posttest. This pattern can be explained by speculating how engaging each of the phases was. The productive failure phase may have been the most engaging both because of the collaboration and because of the mental effort needed during problem solving. The lesson, on the other hand, appeared to be less engaging, despite the fact we tried to make it as engaging as possible by adding in animated slides and dialogue style question and answers that prior work showed increases learning (Driscoll et al., 2003). The decrease in boredom after the posttest suggests participants found it to be more engaging than the lesson. This is not surprising given that a test situation involves active problem solving, while the lesson required passive listening.

The coded verbal transcripts and the experience sampling questionnaires were both used to measure emotions. While the measures did have some overlap, they each

also each measured unique emotions. Recall we coded for three emotions in the verbal protocols: confusion, uncertainty, and frustration. For the experience sampling, we measured five emotions: anxiety, boredom, enjoyment, frustration, and confusion. While confusion was measured in both cases, there are two differences between these instruments: (1) uncertainty is not included in the experience sampling questionnaire, (2) enjoyment, boredom and anxiety are not included in the coding scheme. Uncertainty was not included in the experience sampling because when it is not specifically tied to ideation, uncertainty and confusion are more difficult to distinguish. Anxiety, boredom, and enjoyment were also excluded from the coding scheme. We did not code for enjoyment because as noted above we wanted to focus on negative emotions. While we speculated that anxiety and boredom could occur during productive failure as they have been found in other learning contexts (Pekrun et al., 2011), they would be difficult to distinguish in the verbal protocols. We therefore used self-reports in the experience sampling questionnaire to measure these emotions.

The emotions measured during productive failure were state emotions; this was also the case for the emotions measured by experience sampling. In addition to measures of state emotions, we also measured trait emotions at the beginning of the experiment using the Academic Emotions Questionnaire (AEQ). We originally intended to use these emotions as a baseline comparison for the state emotions measured. Upon further reflection, we realized that the two measures were distinct (trait emotions and state emotions), because the trait emotions assessed participants' general feelings about academic contexts as opposed to the state emotions, which measured participants' feelings at a specific moment. The levels of AEQ emotions reported in our study were

similar to ones reported by Pekrun et al. (2011) who used the AEQ with a similar population. Most of the emotions fell between just below the median of the scale (2.5) to slightly above the median (3.5) with pride being slightly more prevalent and hopelessness being slightly less prevalent in our study.

In addition to emotions, we measured the presence of related academic states, including self-efficacy, curiosity, and intolerance for uncertainty. For self-efficacy, our results were similar to a study that validated the questionnaire with a similar population (Imam, 2007). For the curiosity questionnaire, we did not find comparable studies to compare our results to; other studies used a scaled version of the instrument with 10 elements and so we could not compare our results (Ruiz-Alfonso & Leon, 2019). Finally, the level of intolerance for uncertainty in our study was higher (by 20 points) than what was found in a study validating the questionnaire in a similar population (Buhr & Dugas, 2002) – the cause of this difference is unclear.

## **5.2 Learning and Emotion in Productive Failure**

Two pairs solved the problem during the productive failure phase (the remaining 22 pairs did not). While being able to solve the problem contradicts the main premise of productive failure, namely that students should fail to find a solution, these participants still had to fail during the process of searching for the solution. Thus, they were able to recognize knowledge gaps and activate prior knowledge in order to solve the problem. For example, PF117 realized that a necessary part of the solution was to swap values in the list. To do this, the pair created a temporary variable that allowed them to store the value of the current smallest item. They then thought that they could insert that value and delete the smallest value from the list. There were however, two mistakes with their plan:

(1) they were attempting to insert a value at a position that was not the position of the smallest number (e.g. they wanted to insert the value at position one when position two was the position of the smallest element), and (2) they attempted to insert and delete inside of their 'if' statement. They did eventually realize that insertion and deletion were necessary but that they had to be done outside of the 'if' and that the value being inserted had to be represented by the position of the current smallest and not the value itself. This shows how the pair discovered a knowledge gap and then used their prior knowledge to help overcome it.

PF108, the second pair that produced a solution eventually, also made a mistake with their 'if' when they tried to compare a single number to an item at some position. Their first fix was to swap the two (comparing an item at some position to a set item). While this was an improvement, they were still considering the elements themselves and not the positions, which is a critical concept of sorting. After a while, they did realize that position was important. Thus, even though they ultimately solved the problem, the failure they experienced allowed them to become aware of knowledge gaps and learn to recognize deeper features of the problem space.

As far as general learning, we found a significant increase in procedural knowledge between pretest and posttest. This was unusual for a productive failure paradigm because in prior work it has not affected procedural learning (Kapur, 2008, 2012). The learning in the present study may be attributable to the fact that the instructional materials during the experiment covered some of the concepts needed to solve the problems on the tests; moreover, our materials included tutorials that prior studies did not.

For conceptual learning, participants scored lower on the posttest conceptual questions than the procedural questions and in general the scores for the conceptual parts of the test (labelled recall and conceptual in Figure 4.4) were low. When we compared our results with Kapur's work (2008, 2009, 2012), we found students in these prior studies scored higher on measures of conceptual knowledge. For example, Kapur (2012) showed that students' mean posttest score for productive failure was 87.6% whereas our mean posttest performance for conceptual learning was 47%. One possible reason for this difference could have been that students in our study did not have adequate prior knowledge. This explanation, however, is unlikely as sorting is a familiar 'everyday' type task, and the lesson did provide the necessary conceptual foundations for the posttest's questions. Eight out of the 10 posttest conceptual questions were created based on content from the lesson with the other two being transfer questions that required participants to apply concepts to novel situations. Moreover, productive failure is designed to increase conceptual knowledge because it prepares students to learn concepts presented in the lesson by asking them to solve problems based on the concepts. Another explanation for the relatively low conceptual scores could be that our consolidation phase was conducted differently than in a standard productive failure paradigm. In the standard paradigm, consolidation consists of a whole-group discussion where the instructor explicitly goes over errors in solution methods and discusses how to solve the problem. For our consolidation phase, the lesson was delivered via a video that participants could not pause or rewind. Though the lesson was still designed based on the guidelines given by Kapur and Bielaczyc (2012), i.e. common errors were discussed and refuted and the canonical solution was provided, perhaps not allowing the participants to stop the video and discuss

its contents with each other reduced the conceptual scores. A final possibility for the cause of the difference is that programming is a more challenging domain than the domains used in the prior studies.

Prior work shows that there is a relationship between emotions and learning (Pekrun, Elliot & Maier, 2009). In contrast, when we examined the relations between emotions and posttest scores, we found that the state emotions were not predictive of posttest, for reasons we speculate on below. The only exception to this was that confusion measured directly after the posttest negatively predicted posttest outcomes. This is expected given that the more confused one feels about a subject, the lower their performance on a test concerning that subject will be. Contrary to this, D’Mello et al. (2014) found the opposite result: the more confused students were, the better their learning outcomes were. However, D’Mello induced confusion during the problem-solving phase and that may have led students to reflect more on their knowledge gaps and potential errors, which may have fostered learning. In the present study, the confusion measured after the posttest would not have any reason to positively correlate with learning because it was confusion to do with the posttest measured after the test.

One possible explanation for why the emotions measured immediately after the instructional activities (e.g., productive failure) did not correlate with posttest scores is because our participants experienced other emotions during these activities, but these changed by the time the experience sampling was collected. For instance, the emotions participants felt right after productive failure may have been different from the emotions they experienced during productive failure. Thus, it is not entirely unexpected that those measures did not correlate.

In addition to finding no evidence of relations between state emotions and learning, we also found no evidence of relations between trait emotions and learning. This is also contrary to previous work. For example, using the same scale and with a similar population, Mega, Ronconi and De Beni (2013) showed that positively-valenced emotions correlated positively with learning while negatively-valenced emotions correlated negatively with learning. There is, however, a plausible explanation for our lack of results: the relations mentioned in the prior studies had small effects (all correlation coefficients  $r < .3$ ). Thus, we would have needed a larger sample to find significant relations between trait emotions and learning.

As far as intolerance for uncertainty, prior studies found that higher intolerance for uncertainty was associated with reduced learning (Keshavarz & Assar, 2009). We did not find evidence for any relation. This again could be due to the modest sample size in the present study as prior work that did find an association between intolerance for uncertainty and learning had a sample size four times larger than ours.

Finally, we anticipated that the number of ideas expressed by participants in the verbal protocols would correlate with their posttest outcomes. This conjecture was based by work by Chi, Lam, and Muldner (2014) who found that the more substantive contributions students made, the more they learned. Substantive contributions were defined as "...a meaningful contribution to an ongoing activity, such as problem solving, or a relevant response to a tutor's explanations." (Chi, Lam, and Muldner, 2014, p. 78) and so are similar to ideas in the present study. We did not however, find evidence for this relationship in our study. One possible reason for this is that the ideas expressed by participants could have been mostly related to the procedural aspects of the problem as

opposed to the conceptual aspects. For example, an idea like ‘we need a nested loop’ is conceptually relevant; one of the concepts behind sorting is that the program needs to have an inner loop go through the list to find the position of the smallest value while the outer loop inserts values into the new list and deletes them from the old list. But an idea such as ‘we need to drag a loop block in that is set to run 10 times’ is procedural. If this is the case, then it provides an explanation for why the number of ideas did not correlate with posttest scores because the majority of the posttest was focused on conceptual knowledge rather than procedural.

### **5.3 Limitations and Future Work**

One of the unanticipated effects of the COVID-19 pandemic was that the study had to be moved online, which caused unanticipated delays. One limitation that arose as a result of this is that we did not have the resources to train a second coder for the verbal transcripts. We aimed to compensate for this by having the thesis supervisor and author independently code and then compare notes on the first three transcripts in order to refine the coding scheme. Even so, for future work, having more than one coder and measuring inter-rater reliability should be included in order to maximize the validity of the coding.

Changing to an online study also meant that we had to reconfigure our instructional materials for an online medium and consider how collaboration would occur in an online setting. While students do collaborate online, the collaboration might be richer if done in person rather than virtually. Prior productive failure studies have included in-person collaboration, and so our results have to be interpreted with the caveat that online collaboration might be different than in-person collaboration. An interesting future avenue would be to analyze how much collaboration was occurring, which would

require further coding of the verbal protocols. Analyzing collaboration was beyond the scope of the present thesis and so we do not know how much collaboration occurred. If the amount was low, another avenue for future work would be to include ways to induce collaboration other than mentioning it in the task instructions.

Another future work direction relates to the lesson. As mentioned above, the design of the lesson follows productive failure guidelines, including mentioning misconceptions about sorting and refuting them. This was done in accordance with both Kapur & Bielaczyc's (2012) and Tawfik, Rong & Choi's (2015) guidelines that demonstrate the importance of including contrasting cases in productive failure. In this case, the common misconceptions are the single element on which a contrasting case would be based. For example, the difference between a solution with or without a nested loop. As for the presentation of the lesson, prior productive failure work was predominantly carried out in the classroom and a teacher delivered the lesson to the entire class at once. In contrast, in the current study, participants were run individually and so we used a pre-recorded lesson to ensure that all participants received the same instructional content. For future work, we plan to explore how to make the format of the lesson more engaging for students. One possible way to accomplish this is to embed questions in the lesson by the instructor to the audience, including asking participants to pause the video to reflect on the answer.

Another potential limitation pertains to timing. The experiment was long (around two hours) despite the fact that we aimed to streamline as much as possible. It is possible that some fatigue occurred over the course of the experiment. Moreover, the online context may have exacerbated this fatigue. For most of the experiment, participants'

webcams were turned off to preserve privacy (e.g., during the recording of the productive failure phase) and this may have adversely affected participants' engagement in the experiment. With no direct oversight by the experimenter, participants may have been distracted by other activities while they were supposed to be engaged with the experiment. One possible solution to this issue and the lesson delivery limitation mentioned above would involve running the experiment in a lab or classroom setting.

The present method used a single condition study. While we felt that this was appropriate given that there has been no prior investigation of these relationships, it also meant that we had trouble evaluating the impacts of emotion on productive failure learning; we had no way of comparing the effects of emotions in productive failure to another paradigm in order to find how large the effect was. A two-condition study would have allowed us to contrast productive failure with a regular learning paradigm, including the role of emotion in each. In the future, we plan to run a two-condition study where students receive either a productive failure or a standard order learning intervention where a concept is taught first and then students are given problems to solve (direct instruction). This condition has been used as a comparison for productive failure in the past (Kapur, 2008, 2012) and has been shown to engender less learning than productive failure. The study would run almost identically to the current one except that there would be two conditions. This would allow us to examine the relationship between emotions and learning across condition and to investigate whether condition influences if/how much emotions are predictive of learning.

We did not account for the dyadic nature of our data. The analyses performed assumed that each participants' emotional state was unaffected by their partner's and/or

the collaboration itself. One solution to this would be to use a multilevel model that accounts for both the individual levels of emotions as well as the effect of collaboration.

Another limitation was the lack of a delayed posttest. While it was impractical for this particular experiment due to time constraints, a delayed posttest should be included in future work as it would allow us to measure participants' retention, especially of the conceptual information, after the experiment was finished.

Yet another direction for future work would be to include positive emotions in the coding scheme. Based on our original conjecture that the levels of negative emotions should be higher than the levels of positive emotions, focusing on negatives over positives was warranted. In light of our finding that the occurrence of expression of negative emotions in productive failure was lower than expected, it would be interesting to examine whether positive emotions were more or less prevalent.

#### **5.4 Conclusion**

In the present work, the presence of emotions and how they changed over various instructional activities was examined in a productive failure paradigm, along with the relation between productive failure learning and emotions. As expected, emotions were present in productive failure and the paradigm did impact emotion (e.g., anxiety was high after the productive failure phase, then dropped after the lesson and increased slightly after the posttest). However, we did not find evidence for the relationship between emotions and learning predicted by prior studies. One implication is that productive failure did not induce high levels of negative emotions despite the fact that it involves failure, but this finding needs further investigation with experimental studies.

## References

- Baker, R. S. J. d., D'Mello, S. K., Rodrigo, M. T. & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, 68(4), 223-241. <https://doi.org/10.1016/j.ijhcs.2009.12.003>.
- Berkley University, (N/A) Snap! Build your own blocks. Retrieved from: <http://snap.berkeley.edu/>
- Borracci, G., Gauthier, E., Jennings, J., Sale, K., & Muldner, K. (2020). The Effect of Assistance on Learning and Affect in an Algebra Tutor. *Journal of Educational Computing Research*, 57(8), 2032–2052. <https://doi.org/10.1177/0735633118822103>
- Buhr, K., & Dugas, M.J. (2002) The intolerance of uncertainty scale: psychometric properties of the English version. *Behavior Research and Therapy*, 40(8), 931-945. [https://doi.org/10.1016/S0005-7967\(01\)00092-4](https://doi.org/10.1016/S0005-7967(01)00092-4).
- Chapelle, C. and Roberts, C. (1986). Ambiguity Tolerance and Field Independence as Predictors of Proficiency in English as a Second Language. *Language Learning*, 36, 27-45. [10.1111/j.1467-1770.1986.tb00367.x](https://doi.org/10.1111/j.1467-1770.1986.tb00367.x)
- Costa, J.M. & Miranda, G.L. (2017). Relation between Alice software and programming learning: A systematic review of the literature and meta-analysis. *Br J Educ Technol*, 48, 1464-1474. [10.1111/bjet.12496](https://doi.org/10.1111/bjet.12496)

- Craig, S., Graesser, A., Sullins, J. & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media*, 29(3), 241-250. [10.1080/1358165042000283101](https://doi.org/10.1080/1358165042000283101)
- Darabi, A., Arrington, T.L. & Sayilir, E. (2018) Learning from failure: a meta-analysis of the empirical studies. *Education Tech Research Dev*, 66, 1101-1118. <https://doi.org/10.1007/s11423-018-9579-9>
- D'Mello, S. K., Graesser, A. C., Lehman, B. & Pekrun, R. (2014) Confusion can be beneficial for learning, *Learning and Instruction*, 29, 153-170. <https://doi.org/10.1016/j.learninstruc.2012.05.003>
- D'Mello, S. K., & Graesser, A. C. (2014). Confusion. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *Educational psychology handbook series. International handbook of emotions in education* (pp. 289–310). Routledge/Taylor & Francis Group.
- Driscoll, D. M., Craig, S. D., Gholson, B., Ventura, M., Hu, X., & Graesser, A. C. (2003). Vicarious Learning: Effects of Overhearing Dialog and Monologue-like Discourse in a Virtual Tutoring Session. *Journal of Educational Computing Research*, 29(4), 431–450. <https://doi.org/10.2190/Q8CM-FH7L-6HJU-DT9W>
- Falout, J., Elwood, J., & Hood, M. (2009). Demotivation: Affective states and learning outcomes. *System*, 37(3), 403-417. doi:10.1016/j.system.2009.03.004
- Fang He, V., Sirén, C., Singh, S., Solomon, G., & von Krogh, G. (2018). Keep Calm and Carry On: Emotion Regulation in Entrepreneurs' Learning from Failure. *Entrepreneurship Theory and Practice*, 42(4), 605–630. <https://doi.org/10.1177/1042258718783428>

- Freeston, M. H., Rhéaume, J., Letarte, H., Dugas, M. J. & Ladouceur, R. (1994). Why do people worry? *Personality and Individual Differences*, 17(6), 791-802.  
[https://doi.org/10.1016/0191-8869\(94\)90048-5](https://doi.org/10.1016/0191-8869(94)90048-5)
- Gadgil, S., Nokes-Malach, T., & Chi, M. (2012). Effectiveness of holistic mental model confrontation in driving conceptual change. *Learning and Instruction*, 22(1), 47-61. 10.1016/j.learninstruc.2011.06.002
- Goetz, T., & Hall, N. C. (2013). Emotion and Achievement in the Classroom. In *International guide to student achievement* (pp. 192-195). New York, NY: Routledge.
- Goetz, T., Sticca, F., Pekrun, R., Murayama, K. & Elliot, A. J. (2015). Intraindividual relations between achievement goals and discrete achievement emotions: An experience sampling approach, *Learning and Instruction*, 41, 115-125.  
<https://doi.org/10.1016/j.learninstruc.2015.10.007>
- Imam, S. S. (2007). Sherer et al. General Self-Efficacy Scale: Dimensionality, Internal Consistency, and Temporal Stability. *Redesigning Pedagogy: Culture, Knowledge and Understanding*, Singapore.  
[http://www.bwgriffin.com/gsu/courses/edur9131/content/3\\_FA\\_general\\_efficiency.p.pdf](http://www.bwgriffin.com/gsu/courses/edur9131/content/3_FA_general_efficiency.p.pdf)
- Jeronimus, B. F., Laucelle, O. M. (2017). Frustration. In L. Zeigler-Hill V., Shackelford T. (eds) *Encyclopedia of Personality and Individual Differences*. Springer, Cham.  
[https://doi.org/10.1007/978-3-319-28099-8\\_815-1](https://doi.org/10.1007/978-3-319-28099-8_815-1)
- Kapur, M., (2008), Productive Failure, *Cognition and Instruction*, 26(3), 379-424. [10.1080/07370000802212669](https://doi.org/10.1080/07370000802212669)

- Kapur, M. (2011). A further study of productive failure in mathematical problem solving: Unpacking the design components. *Instructional Science*, 39(4), 561–579.  
[10.1007/s11251-010-9144-3](https://doi.org/10.1007/s11251-010-9144-3)
- Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science*, 40(4), 651-672. <http://dx.doi.org/10.1007/s11251-012-9209-6>
- Kapur, M. (2014). Comparing Learning From Productive Failure and Vicarious Failure. *Journal of the Learning Sciences*, 23(4), 651-677.  
[10.1080/10508406.2013.819000](https://doi.org/10.1080/10508406.2013.819000)
- Kapur, M. (2016). Examining Productive Failure, Productive Success, Unproductive Failure, and Unproductive Success in Learning. *Educational Psychologist*, 51(2), 289-299. [10.1080/00461520.2016.1155457](https://doi.org/10.1080/00461520.2016.1155457)
- Kapur, M. & Bielaczyc, K. (2012). Designing for Productive Failure. *Journal of the Learning Sciences*, 21(1), 45-83. [10.1080/10508406.2011.591717](https://doi.org/10.1080/10508406.2011.591717)
- Kapur, M., Kinzer, C.K. (2009). Productive failure in CSCL groups. *Computer Supported Learning* 4, 21-46. <https://doi.org/10.1007/s11412-008-9059-z>
- Keshavarz, M., Assar, M. (2011). Reading Comprehension Ability and Metacognitive Awareness of Reading Strategies among High, Mid and Low Ambiguity Tolerance EAP Students. *Iranian Journal of Applied Language Studies*, 1(2), 71-108. doi: [10.22111/ijals.2011.53](https://doi.org/10.22111/ijals.2011.53)
- Kim, C. & Pekrun, R. (2014). Emotions and Motivation in Learning and Performance. *Handbook of Research on Educational Communications and Technology*, 65-75.  
doi:[10.1007/978-1-4614-3185-5\\_6](https://doi.org/10.1007/978-1-4614-3185-5_6).

- Kort B., Reilly, R. & Picard, R. W. (2001). External representation of learning process and domain knowledge: affective state as a determinate of its structure and function, In: *Proceedings of the Artificial Intelligence in Education Workshops* (San Antonio, TX, AIED), 64-69. URL: [https://www.researchgate.net/publication/2867488\\_External\\_Representation\\_of\\_Learning\\_Process\\_and\\_Domain\\_Knowledge\\_Affective\\_State\\_as\\_a\\_Determinate\\_of\\_its\\_Structure\\_and\\_Function](https://www.researchgate.net/publication/2867488_External_Representation_of_Learning_Process_and_Domain_Knowledge_Affective_State_as_a_Determinate_of_its_Structure_and_Function)
- Lamnina, M., & Chase, C.C. (2019). Developing a thirst for knowledge: How uncertainty in the classroom influences curiosity, affect, learning, and transfer. *Contemporary Educational Psychology*, 59. <https://doi.org/10.1016/j.cedpsych.2019.101785>.
- Lane, A. M., Whyte, G. P., Terry, P. C., & Nevill, A. M. (2005). Mood, self-set goals and examination performance: The moderating effect of depressed mood. *Personality and Individual Differences*, 39(1), 143–153. <https://doi.org/10.1016/j.paid.2004.12.015>
- Litman, J., & Spielberger, C. (2003). Measuring Epistemic Curiosity and Its Diverse and Specific Components. *Journal of Personality Assessment*, 80:1, 75-86. 10.1207/S15327752JPA8001\_16
- Lix, L. M., Keselman, J. C., & Keselman, H. J. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test. *Review of Educational Research*, 66(4), 579–619. <https://doi.org/10.2307/1170654>

- Loibl, K., & Rummel, N. (2014). Knowing what you don't know makes failure productive. *Learning and Instruction*, 34, 74-85.  
10.1016/j.learninstruc.2014.08.004
- Loibl, K., Roll, I. & Rummel, N. (2017). Towards a Theory of When and How Problem Solving Followed by Instruction Supports Learning. *Educ Psychol Rev*, 29, 693-715. <https://doi.org/10.1007/s10648-016-9379-x>
- MacIntyre, P., & Vincze, L. (2017). Positive and negative emotions underlie motivation for L2 learning. *Studies in Second Language Learning and Teaching*, 7(1), 61-88.  
<https://doi.org/10.14746/ssllt.2017.7.1.4>
- Mazziotti, C., Loibl, K., & Rummel, N. (2015). Collaborative or Individual Learning within Productive Failure: Does the Social Form of Learning Make a Difference? In Lindwall, O., Häkkinen, P., Koschman, T. Tchounikine, P. Ludvigsen, S. (Eds.) (2015). *Exploring the Material Conditions of Learning: The Computer Supported Collaborative Learning (CSCL) Conference 2015, Volume 2*. Gothenburg, Sweden: The International Society of the Learning Sciences. URL: <https://repository.isls.org/handle/1/468>
- Mega, C., Ronconi, L., & De Beni, R. (2014). What makes a good student? How emotions, self-regulated learning, and motivation contribute to academic achievement. *Journal of Educational Psychology*, 106(1), 121–131. <https://doi.org/10.1037/a0033546>
- Muldner, K., Lam, R., & Chi, M. T. H. (2014). Comparing learning from observing and from human tutoring. *Journal of Educational Psychology*, 106(1), 69–85. <https://doi.org/10.1037/a0034448>

- Muldner K., Wixon M., Rai D., Burleson W., Woolf B., Arroyo I. (2015) Exploring the Impact of a Learning Dashboard on Student Affect. In: Conati C., Heffernan N., Mitrovic A., Verdejo M. (eds) Artificial Intelligence in Education. AIED 2015. Lecture Notes in Computer Science, vol 9112. Springer, Cham.  
[https://doi.org/10.1007/978-3-319-19773-9\\_31](https://doi.org/10.1007/978-3-319-19773-9_31)
- Nummenmaa, L., & Niemi, P. (2004). Inducing affective states with success-failure manipulations: A meta-analysis. *Emotion*, 4(2), 207-214.  
<http://dx.doi.org.proxy.library.carleton.ca/10.1037/1528-3542.4.2.207>
- Oudeyer, P-Y., Gottlieb, J. & Lopes, M. (2016). Intrinsic motivation, curiosity and learning: theory and applications in educational technologies. *Progress in brain research, Elsevier*, 229, 257-28. ff10.1016/bs.pbr.2016.05.005ff. ffhal-01404278f
- Pataranutaporn, V., & Liu, Z. (2013). Sequences of Frustration and Confusion, and Learning. Retrieved from <https://digitalcommons.wpi.edu/iqp-all/3296>
- Pekrun, R. (2006). The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educ Psychol Rev*, 18, 315–341. <https://doi.org/10.1007/s10648-006-9029-9>
- Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic Emotions in Students' Self-Regulated Learning and Achievement: A Program of Qualitative and Quantitative Research. *Educational Psychologist*, 37(2), 91-105. [10.1207/S15326985EP3702\\_4](https://doi.org/10.1207/S15326985EP3702_4)
- Pekrun, R., Elliot, A. J., & Maier, M. A. (2009). Achievement goals and achievement emotions: Testing a model of their joint relations with academic

- performance. *Journal of Educational Psychology*, 101(1), 115–135. <https://doi.org/10.1037/a0013383>
- Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P. & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ), *Contemporary Educational Psychology*, 36(1), 36-48. <https://doi.org/10.1016/j.cedpsych.2010.10.002>
- Pekrun, R., & Linnenbrink-Garcia, L. (2012). *Academic emotions and student engagement*. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (p. 259–282). Springer Science + Business Media. [https://doi.org/10.1007/978-1-4614-2018-7\\_12](https://doi.org/10.1007/978-1-4614-2018-7_12)
- Pekrun, R. & Stephens, E.J. (2010). Achievement Emotions: A Control-Value Approach. *Social and Personality Psychology Compass*, 4, 238-255. [10.1111/j.1751-9004.2010.00259.x](https://doi.org/10.1111/j.1751-9004.2010.00259.x)
- Penrod, J. (2001). Refinement of the concept of uncertainty. *Journal of Advanced Nursing*, 34, 238-245. [10.1046/j.1365-2648.2001.01750.x](https://doi.org/10.1046/j.1365-2648.2001.01750.x)
- Piaget, J. (1977). *The development of thought: Equilibration of cognitive structures*. (Trans A. Rosin). Viking.
- Pietarinen, T., Vauras, M., Laakkonen, E., Kinnunen, R., & Volet, S. (2018). High school students' perceptions of affect and collaboration during virtual science inquiry learning. *J Comput Assist Learn*, 35, 334– 348. <https://doi.org/10.1111/jcal.12334>
- Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal*

of *Educational Psychology*, 93(2), 346–362. <https://doi.org/10.1037/0022-0663.93.2.346>

Ruiz-Alfonso, Z., & León, J. (2019). Teaching quality: relationships between passion, deep strategy to learn, and epistemic curiosity. *School Effectiveness and School Improvement*, 30(2), 212–230. <https://doi.org/10.1080/09243453.2018.1562944>

Schank, R. C. (1999). Dynamic memory revisited. *Cambridge University Press*. <https://doi.org/10.1017/CBO9780511527920>

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44(4), 695–729. <https://doi.org/10.1177/0539018405058216>

Sherer, M., Maddux, J. E., Mercandante, B., Prentice-Dunn, S., Jacobs, B., & Rogers, R. W. (1982). The Self-Efficacy Scale: Construction and Validation. *Psychological Reports*, 51(2), 663-671. <https://doi.org/10.2466/pr0.1982.51.2.663>

Sinha, T., & Kapur, M. (2019). When Productive Failure Fails. In Proceedings of the Annual Meeting of the Cognitive Science Society

Staus, N. L. and Falk, J. H. (2017). The Role of Emotion in Informal Science Learning: Testing an Exploratory Model. *Mind, Brain, and Education*, 11, 45-53. [10.1111/mbe.12139](https://doi.org/10.1111/mbe.12139)

Tawfik, A.A, Rong, H. & Choi, I. (2015). Failing to learn: towards a unified design approach for failure-based learning. *Education Tech Research Dev*, 63, 975-994. <https://doi.org/10.1007/s11423-015-9399-0>

Toh, P. L.L., & Kapur, M., (2017). Is having more prerequisite knowledge better for learning from productive failure?. *Instructional Science*, 45, 377-394. [10.1007/s11251-016-9402-0](https://doi.org/10.1007/s11251-016-9402-0)

- VanLehn, K. (1988). Toward a theory of impasse-driven learning. In Mandl, D. H. & Lesgold, D. A. (Eds.) *Learning Issues for Intelligent Tutoring Systems* (pp. 19–41). Springer US. Retrieved from [http://link.springer.com/chapter/10.1007/978-1-4684-6350-7\\_2](http://link.springer.com/chapter/10.1007/978-1-4684-6350-7_2)
- Vatcheva, K. P., Lee, M., McCormick, J. B., & Rahbar, M. H. (2016). Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies. *Epidemiology*, 6(2), 227. <https://doi.org/10.4172/2161-1165.1000227>
- Villavicencio, F.T. and Bernardo, A.B.I. (2013). Positive academic emotions moderate the relationship between self-regulation and academic achievement. *British Journal of Educational Psychology*, 83, 329-340. [10.1111/j.2044-8279.2012.02064.x](https://doi.org/10.1111/j.2044-8279.2012.02064.x)
- Wundt, W. (1905) *Grundzüge der physiologischen Psychologie*. Leipzig: Engelmann.

# Appendices

## Appendix A

### A.1 Pretest

Participant ID \_\_\_\_\_

Date \_\_\_\_\_

```
set position to 1
set ListA to list 3 5 2 1 4
set result to 0
repeat (length of ListA)
  change result by item position of ListA
  change position by 1
say result
```

Given the program on the left, show what it will print [show your work].

2. Given a list of numbers write a program that inserts scores greater than 20 but less than 30 into a new list.

```
list 21 5 41 22 29 19 50 2 27 6
```

Your code should work with a list of any length. You can write your program in pseudo code (plain English).

## A.2 Posttest

Participant ID \_\_\_\_\_

Date \_\_\_\_\_

1. Given the program below, show what it will print [show your work].

```
set position to 1
set ListA to list 3 5 2 1 4
set result to 0
repeat length of ListA
  change result by item position of ListA
  change position by 1
say result
```

2. Given a list of numbers write a program that inserts scores greater than 20 but less than 30 into a new list.

```
list 21 5 41 22 29 19 50 2 27 6
```

Your code should work with a list of any length. You can write your program in pseudo code (plain English).

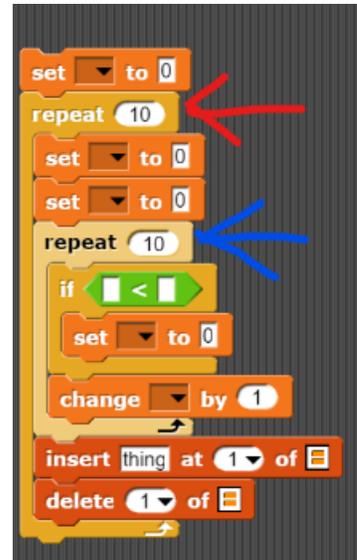
3. Given a list of numbers, write a general *algorithm* to sort the list from greatest to smallest.

Participant id: \_\_\_\_\_

Date \_\_\_\_\_

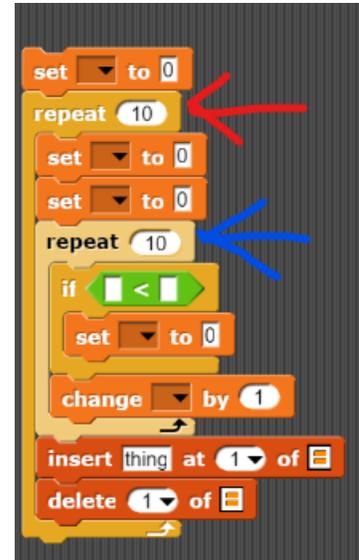
All questions relate to the sorting program lesson. Please answer them in plain English – provide as much detail as possible. A skeleton of a snap sorting program is shown below.

4. The sorting lesson included four “big ideas” you need to sort. List these.



5. In the sorting program, we remove the smallest score from the original list for each iteration of the outer loop (outer loop is indicated by red arrow, see skeleton code on the right). Why is this “removal” necessary?
6. What is the purpose of the outer loop – see red arrow, above right (i.e., what function is it performing that helps the program sort scores)?
7. What is the purpose of the inner loop – see blue arrow, above right ? (i.e., what function is it performing that helps the program sort scores?)

8. Our original list had 10 numbers in it. Suppose we stop the program after only 3 iterations of the outer loop (see red arrow, right). What would the results list contain?



9. Why did the program need to keep track of positions of scores instead of just the scores themselves?
10. Why does the program need to set minPosition and position to 1 each time the outer loop cycles around (see red arrow showing outer loop, snap program top right) – why not just set them to 1 once outside the outer loop?

### A.3 Grading Scheme

#### **Pretest:**

Question 1: 1 or 0 (correct answer is 15)

Question 2: out of 2

0.5 for loop

0.5 for conditional

0.5 for insert

0.5 for some mention of new list

#### **Posttest part 1 (Question 1 &2):**

Identical scoring to Pretest

#### **Posttest part 2 (Question 3):**

0.5 for mentioning creation of new variable(s)

0.5 for setting positions to 1 at start of outer loop

0.5 for mentioning 1 loop → 1 point for mentioning 2

0.5 for having an if statement

0.5 if the “if” is correctly phrased.

0.5 for changing position by 1 after the “if”

0.5 for mentioning a new list for sorting

0.5 for mentioning insertion of smallest element into new list

0.5 for mentioning deletion of element from original list

#### **Posttest part 3 (Questions 4-10):**

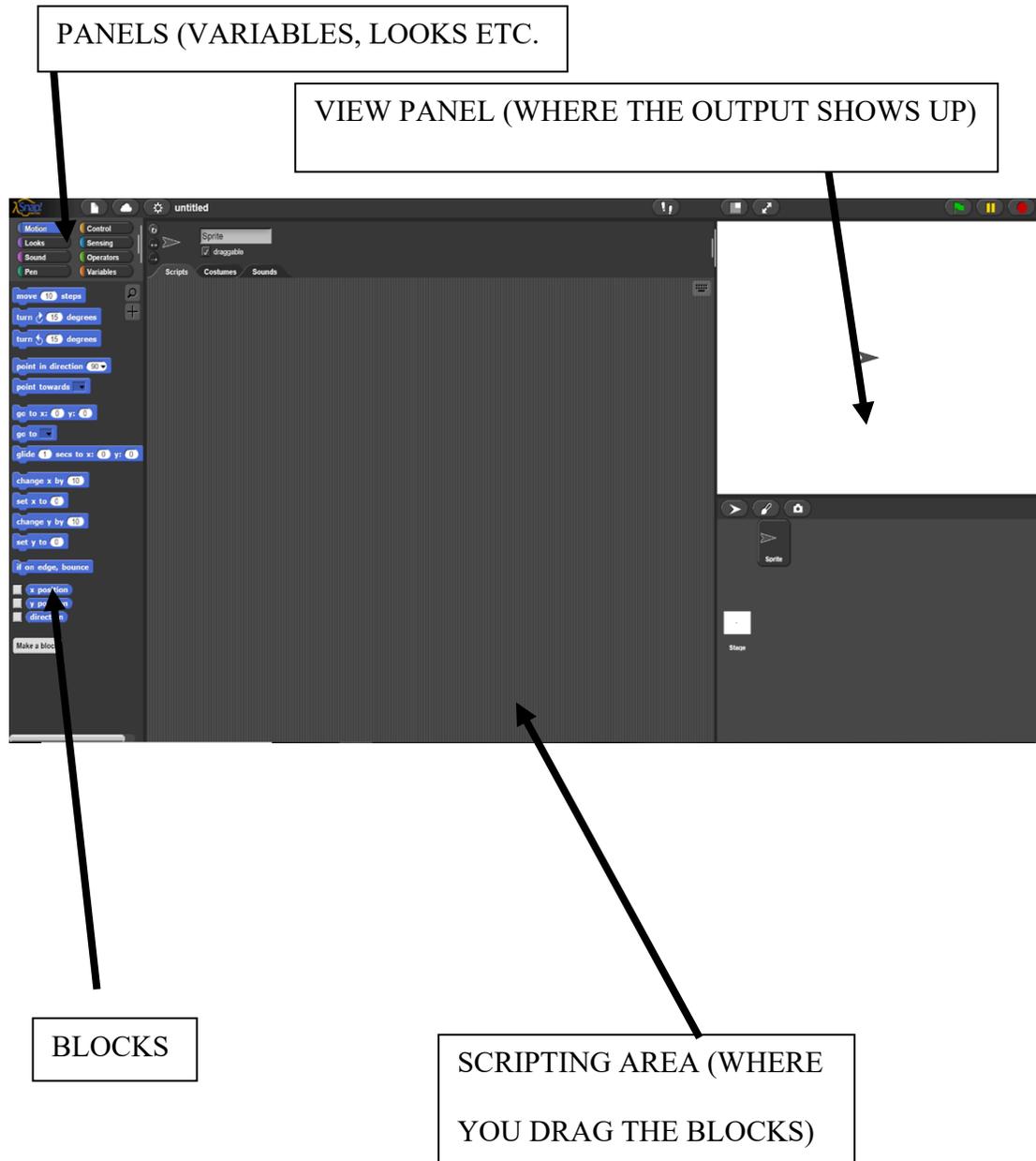
Questions 4: worth 4 points (1 point for mentioning each of the big ideas: (1) new list, (2) deleting from old list, (3) needing 2 loops, and (4) finding position of smallest element [not smallest value]).

Questions 5-11: worth 1 point each

Note: half points can be awarded if a participant is on the right track for a correct answer (this means that they have mentioned an element of a correct answer or are missing one component of a correct answer).

## Appendix B

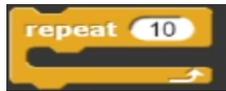
### B.1 SNAP! Interactive Cheat Sheet



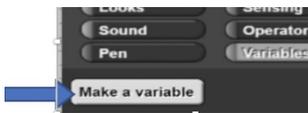
## B.2 SNAP! Block Definitions and FAQ Cheat Sheet



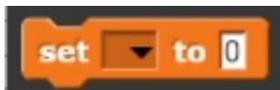
This is a conditional block. It asks you to fill in a condition to meet (ex. is  $X > Y$ ) and if the condition is met, it will execute whatever is placed inside the block (ex. add  $X+2Y$ ). This type of block is used to ensure that only the parts of our data we want to manipulate get used.



This is a looping block. It allows you to repeat whatever function is placed inside it as many times as you need. The 'blank' part is how many times we want the loop to run. I.e. Repeat '10' will run that loop 10 times.



The button that says 'Make a variable': This button will open a window which will allow you to name your variable. Underneath that are two options 'For all sprites' and 'For this sprite only'. For all exercises today, you can safely leave the selection on 'For all sprites'.



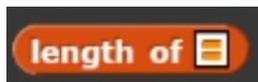
This block is used to set the value of any variable you choose to whatever value you choose. The block itself has a drop-down menu (the little white arrow) which allows you to choose which variable you are setting, and the blank space is where you write the value.



Similar to the last block, this block allows you to change the value of any variable by some increment. The drop-down arrow allows you to select the variable and the blank space is for the value.



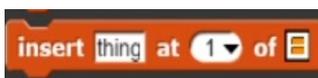
This is the variable block used to make a list. Each white box represents a position in the list which can be filled with a value. Note: you can have a list block with no white boxes which represents an empty list.



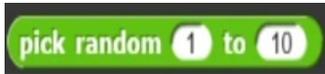
This block stores the length (i.e. how many values there are in) the specified list.



This block lets you select a specific item of a list. For example, if the blank were filled with a 2, the block would look for the second item in list A.



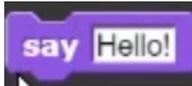
This block inserts a variable (number, word etc.) at the '1' position of whatever list you specify. Remember that both 'thing' and '1' can take variable values.



This block simply picks a random number between the first and second values in the white spaces. Those values can be changed to any numbers.



This is an operator block that simply checks whether some number is less than some other number.



This tells the computer to print out whatever is in the blank space and have the little arrow on the right-hand side of the screen display it in a text bubble.

### Frequently Asked Questions

#### How do I build a SNAP! program?

Drag blocks into the scripting area (the central grey box) to be considered part of a program.

#### How do I get rid of blocks?

Drab them out of the scripting area and drop them over the panel just to the left of it where the variables are.

#### What does it mean to “snap” blocks together?

Blocks can be snapped INTO other blocks by dragging one into the other so that there is NO space between the blocks. A white “bar” will appear if you drag one block close enough to another and they will snap together. Here is what that would look like:



Here is what to blocks look like when they are not snapped together



#### Do all the blocks on a given SNAP! screen have to be snapped together?

Whatever blocks you snap together will run as a single program. There can be more than one “program” on a screen.

#### How do I run a SNAP! program?

You double click the first block in the program (e.g., the “say hello” block in the example above).

### **What if I want to remove some blocks from a program?**

Click on the block you want to remove and drag it away from the series of blocks snapped together. Then do that for each subsequent block.

### **Why won't my variable value change?**

Variable values will not change until you run your program or purposely click on the block that would change them.

## Appendix C

### C.1 Demographic Questionnaire

Participant ID#:

What is your gender?

Male \_\_\_\_\_

Female \_\_\_\_\_

Other \_\_\_\_\_

What is your age?

What is your native language?

What is your Major (preferred area of study if you do not have a major)?

## C.2 Academic Emotions Questionnaire (AEQ)

*Instructions:* Please rate on a scale from 1-5 (1= completely disagree, 3= neutral, 5= completely agree) how true each of these statements are about yourself.

### ***LEARNING-RELATED ENJOYMENT (10) Item***

Question	Response (1-5)
I look forward to studying.	
I enjoy the challenge of learning the material.	
I enjoy acquiring new knowledge.	
I enjoy dealing with the course material.	
Reflecting on my progress in coursework makes me happy.	
I study more than required because I enjoy it so much.	
I am so happy about the progress I made that I am motivated to continue studying.	
Certain subjects are so enjoyable that I am motivated to do extra readings about them.	
When my studies are going well, it gives me a rush.	
I get physically excited when my studies are going well.	

### ***LEARNING-RELATED HOPE (6) Item***

Question	Response (1-5)
I have an optimistic view toward studying.	
I feel confident when studying.	
I feel confident that I will be able to master the material.	
I feel optimistic that I will make good progress at studying.	
The thought of achieving my learning objectives inspires me.	
My sense of confidence motivates me.	

***LEARNING-RELATED PRIDE (6) Item***

Question	Response (1-5)
I'm proud of myself.	
I'm proud of my capacity.	
I think I can be proud of my accomplishments at studying.	
Because I want to be proud of my accomplishments, I am very motivated.	
When I solve a difficult problem in my studying, my heart beats with pride.	
When I excel at my work, I swell with pride.	

***LEARNING-RELATED ANGER (FRUSTRATION) (9) Item***

Question	Response (1-5)
I get angry when I have to study.	
Studying makes me irritated.	
I get angry while studying.	
I'm annoyed that I have to study so much.	
I get annoyed about having to study.	
Because I get so upset over the amount of material, I don't even want to begin studying.	
I get so angry I feel like throwing the textbook out of the window.	
When I sit at my desk for a long time, my irritation makes me restless.	
After extended studying, I'm so angry that I get tense.	

***LEARNING-RELATED ANXIETY (11) Item***

Question	Response (1-5)
When I look at the books I still have to read, I get anxious.	
I get tense and nervous while studying.	
When I can't keep up with my studies it makes me fearful.	
I worry whether I'm able to cope with all my work.	
The subject scares me since I don't fully understand it.	
I worry whether I have properly understood the material.	
I get so nervous that I don't even want to begin to study.	
While studying I feel like distracting myself in order to reduce my anxiety.	
When I have to study, I start to feel queasy.	
As time runs out my heart begins to race.	
Worry about not completing the material makes me sweat.	

***LEARNING-RELATED SHAME (11) Item***

Question	Response (1-5)
I feel ashamed.	
I feel ashamed about my constant procrastination.	
I feel ashamed that I can't absorb the simplest of details.	
I feel ashamed because I am not as adept as others in studying.	
I feel embarrassed about not being able to fully explain the material to others.	
I feel ashamed when I realize that I lack ability.	
My memory gaps embarrass me.	
Because I have had so much troubles with the course material, I avoid discussing it.	
I don't want anybody to know when I haven't been able to understand something.	
When somebody notices how little I understand I avoid eye contact.	
I turn red when I don't know the answer to a question relating to the course material.	

***LEARNING-RELATED HOPELESSNESS (11) Item***

Question	Response (1-5)
I feel hopeless when I think about studying.	
I feel helpless.	
I feel resigned.	
I'm resigned to the fact that I don't have the capacity to master this material.	
After studying I'm resigned to the fact that I haven't got the ability.	
I'm discouraged about the fact that I'll never learn the material.	
I worry because my abilities are not sufficient for my program of studies.	
I feel so helpless that I can't give my studies my full efforts.	
I wish I could quit because I can't cope with it.	
My lack of confidence makes me exhausted before I even start.	
My hopelessness undermines all my energy.	

***LEARNING-RELATED BOREDOM (11) Item***

Question	Response (1-5)
The material bores me to death.	
Studying for my courses bores me.	
Studying is dull and monotonous.	
While studying this boring material, I spend my time thinking of how time stands still.	
The material is so boring that I find myself daydreaming.	
I find my mind wandering while I study.	
Because I'm bored, I have no desire to learn.	
I would rather put off this boring work till tomorrow.	
Because I'm bored, I get tired sitting at my desk.	
The material bores me so much that I feel depleted.	
While studying I seem to drift off because it's so boring.	

### C.3 Self-Efficacy Questionnaire

*Instructions:* For the following items, please rate how much each applies to you on a scale from 1-5 (1= completely disagree, 3= neutral and 5= completely agree).

	Question	Rating (1-5)
1.	When I make plans, I am certain I can make them work	
2.	One of my problems is that I cannot get down to work when I should (R)	
3.	If I can't do a job the first time, I keep trying until I can	
4.	When I set important goals for myself, I rarely achieve them (R)	
5.	I give up on things before completing them (R)	
6.	I avoid facing difficulties (R)	
7.	If something looks too complicated, I will not even bother trying it (R)	
8.	When I have something unpleasant to do, I stick to it until I finish it	
9.	When I decide to do something, I go right to work on it	
10.	When trying to learn something new, I soon give up if I am not initially successful (R)	
11.	When unexpected problems occur, I don't handle them well (R)	
12.	I avoid trying to learn new things when they look too difficult for me (R)	
13.	Failure just makes me try harder	
14.	I feel insecure about my ability to do things (R)	
15.	I am a self-reliant person	
16.	I give up easily (R)	
17.	I do not seem capable of dealing with most problems that come up in life (R)	

## C.4 The Curiosity Questionnaire

*Instructions:* For the following 40 items, please rate how curious you feel about each one on a scale from 1-4 (1= almost never, 2= sometimes, 3= often, 4= almost always).

<u>Question</u>	<u>Rating (1-4)</u>
1. Complex problems and discovering their solutions	
2. Simple explanations of a concept that leave a lot of questions	
3. Complicated machinery/I ask how it works	
4. New solution to difficult problems brings pleasure	
5. I am interested in discovering how things work	
6. Discovering a theory that I would like to test out	
7. Turning new ideas over/thinking about them in different ways	
8. Learn something new/like to find out more about it	
9. Thinking about different answers to the same question	
10. I find it fascinating to learn new information	
11. Read something which puzzles me/keep reading until I understand	
12. Interesting to think about contradicting ideas	
13. New ideas that excite my imagination	
14. Hypothetical situation/like to think about what might happen	
15. Discussing abstract concepts	
16. Incomplete puzzle/try and imagine the final solution	
17. Someone answers a question of mine/I find myself even more inquisitive	
18. It excites me to have a new idea that leads to even more new ideas	
19. I would like to understand how complicated things like computers work	
20. New kind of arithmetic problem/I enjoy imagining solutions	
21. I would enjoy discussing theories about existence with a philosopher	
22. Thinking over new ideas and concepts is fun	
23. I enjoy exploring new ideas	
24. Someone says something ambiguous to me/ I want an explanation	
25. Learning about subjects which are unfamiliar to me	
26. I am interested in how different people would react during a crisis	
27. Something unexpected happens/figure out what caused it	
28. I like to read any magazine that reports new scientific discoveries	
29. Riddle/interested in trying to solve it	
30. I enjoy trying to figure out what led up to important historical events	
31. Words I don't know/look up their meanings	
32. I like to figure out people's motives/behavior	
33. Trying to imagine what people are thinking from their faces	
34. I enjoy trying to understand my feelings	
35. Interested in reading about periods of history that are unfamiliar to me	
36. I like to imagine how a story will end before it's over	
37. When I have a strange dream/I wonder about what it means	
38. Look at a work of art/ I wonder what inspired the artist	
39. I would like to learn more about other cultures	
40. See an advertisement without a caption/wonder what it is about/what it means	

## C.5 The Intolerance for Uncertainty Scale

*Instructions:* For the following 27 items, please rate how characteristic each is of you on a scale from 1-5 (1='not at all characteristic of me' to 5='entirely characteristic of me').

Question	Rating(1-5)
1. Uncertainty stops me from having a strong opinion.	
2. Being uncertain means that a person is disorganized.	
3. Uncertainty makes life intolerable.	
4. It's unfair having no guarantees in life.	
5. My mind can't be relaxed if I don't know what will happen tomorrow.	
6. Uncertainty makes me uneasy, anxious, or stressed.	
7. Unforeseen events upset me greatly.	
8. It frustrates me not having all the information I need.	
9. Uncertainty keeps me from living a full life.	
10. One should always look ahead so as to avoid surprises.	
11. A small unforeseen event can spoil everything, even with the best plan.	
12. When it's time to act, uncertainty paralyzes me.	
13. Being uncertain means that I am not first rate.	
14. When I am uncertain, I can't go forward.	
15. When I am uncertain, I can't function very well.	
16. Unlike me, others seem to know where they are going with their lives.	
17. Uncertainty makes me vulnerable, unhappy, or sad.	
18. I always want to know what the future has in store for me.	
19. I can't stand being taken by surprise.	
20. The smallest doubt can stop me from acting.	
21. I should be able to organize everything in advance.	
22. Being uncertain means that I lack confidence.	
23. I think it's unfair that other people seem to be sure about their future.	
24. Uncertainty keeps me from sleeping soundly.	
25. I must get away from all uncertain situations.	
26. The ambiguities in life stress me.	
27. I can't stand being undecided about my future.	

## C.6 Experience Sampling Questionnaire

*Instructions:* On a scale from 1-5 (1= not at all, 5= extremely) please rate how you are feeling.

1. How anxious are you?
2. How bored are you?
3. How much are you enjoying yourself?
4. How frustrated are you?
5. How confused are you?

## Appendix D

### D.1 Skew and Kurtosis Values for Experience Sampling Measures

Variable	Skewness	Kurtosis
anxiety1	-.15	-1.35
anxiety2	.89	-.13
anxiety3	.16	-1.14
bored1	.81	-.46
bored2	-.43	-.46
bored3	-.03	-.77
enjoyment1	-.04	-.51
enjoyment2	.27	-.54
enjoyment3	.19	-.23
frustration1	-.18	-.98
frustration2	.47	-1.09
frustration3	.15	-1.31
confusion1	-.68	-.81
confusion2	.22	-1.18
confusion3	-.13	-.97

*N* = 48

**D.2 Correlations Between Academic Emotions and Total Posttest Score,  
postTotal, Controlling for Pretest**

Variable	Pearson coefficient (r)	Significance (p)
enjoyment	.02	.88
hope	.09	.54
pride	.02	.88
anger	.13	.40
anxiety	.01	.93
shame	.02	.87
hopelessness	-.13	.41
boredom	.05	.73
self-efficacy	-.02	.91
curiosity	-.03	.87
intolerance for uncertainty	-.24	.11

*N* = 48

## Appendix E



### Informed Consent Form

#### **Name and Contact Information of Researchers:**

Zachary Savelson, Carleton University, Institute of Cognitive Science

Email: zachsavelson@cmail.carleton.ca

Supervisor and Contact Information: Kasia Muldner

Email: kasiamuldner@cunet.carleton.ca

#### **Project Title**

Investigating student learning and emotion during problem-solving activities

#### **Project Sponsor and Funder (if any)**

NSERC Discovery grant

#### **Carleton University Project Clearance**

This study has received clearance by the Carleton University Research Ethics Board B (Clearance #112108).

#### **Invitation**

You are invited to participate in our research study. This study involves working on some online materials with a friend or colleague (you will not be in the same location and thus will communicate using online tools). Thus, to participate you must invite a friend. You both must have *some* programming experience (e.g., from high school, just tinkering, or a basic university class) but not *too much* experience (no more than at one university-level class). You do not have to be good at programming to participate. The information in this form is intended to help you understand what we are asking of you so that you can decide whether you agree to participate in this study. Your participation in this study is voluntary, and a decision not to participate will not be used against you in any way.

As you read this form, and decide whether to participate, please ask all the questions you might have, take whatever time you need, and consult with others as you wish.

#### **What is the purpose of the study?**

The purpose of this study is to examine the relationship between emotions and problem solving. We are using programming as the problem-solving activity.

#### **What will I be asked to do?**

If you agree to take part in the study, we will ask you to fill some questionnaires. One will be on programming – you are not expected to have much background and it is okay if you don't know the answers to it. You will also be asked to fill out a questionnaire that measures mental and emotional states (e.g., school-related emotions, self-efficacy). After

that, you and your partner will work on an activity using Snap! (a programming environment); you will also be given a programming tutorial. You will work virtually but will be able to see each other's work because we will be using an online communication tool (e.g., Zoom). While working with your partner we will audio record the session; we may also record your problem-solving actions in the interface using screen-capture technology (so your face will never be recorded). Please note that audio and screen recording is part of the data collection for this study, and so a condition of consenting to participate. As you work with your partner, you will be asked to self-report on how you are feeling (e.g., frustrated, bored). The audio and screen-recording data will NOT be shared or publicly used and will be destroyed as soon as it is transcribed and analyzed. After the instructional activity, you will fill out a similar set of questionnaires to those at the beginning and will then be given a post-test to assess your learning gains from the activity. The study will take no more than 2 hours.

### **Risks and Inconveniences**

We do not anticipate any risks to participating in this study.

### **Possible Benefits**

You may not receive any direct benefit from your participation in this study. However, your participation may allow researchers to better understand the role of emotions during problem solving.

### **Compensation/Incentives**

As a token of appreciation, you will receive 2% extra credit for CGSC 1001 OR \$25 – the choice is up to you (except that if you choose the bonus credit, you must be enrolled in CGSC1001). If you withdraw before finishing the experiment, the amount of compensation will be prorated by time (e.g. if you withdraw halfway through the study, you will receive half the compensation).

### **No waiver of your rights**

By signing this form, you are not waiving any rights or releasing the researchers from any liability.

### **Withdrawing from the study**

During the study session, you have the right to end your participation for any reason, by stating that you do not want to continue. If you withdraw from the study, all information you have provided will be immediately destroyed (including the audio recording).

Because we do not store any identifying information with the data, once you leave the session, withdrawal is not possible. Should you wish to withdraw, your partner will be given the option to withdraw as well, or to continue with the experiment alone.

### **Confidentiality**

We will treat your personal information as confidential, although absolute privacy cannot be guaranteed. No information that discloses your identity will be released or published. Research records may be accessed by the Carleton University Research Ethics Board in order to ensure continuing ethics compliance.

You will be assigned a code so that your identity will not be directly associated with the data you have provided. However, if you are being granted course credit for taking part in the study, identifying information will be retained using a code until the course credit is granted.

We will password protect any research data that we store or transfer. Research data will only be accessible by the researchers and the research supervisor and will not have any identifiable information. Thus, your name will not appear in any publications or other venues. Once the project is completed, research data will be kept and potentially used for other research projects on this same topic (but as noted above, audio and research data will be destroyed as soon as it is transcribed/analyzed).

**Data Retention**

After the study is completed, your de-identified data will be retained for future research use (note that the audio recordings will be destroyed once they are transcribed by the research team).

**New information during the study**

In the event that any changes could affect your decision to continue participating in this study, you will be promptly informed.

**Ethics review**

This project was reviewed and cleared by the Carleton University Research Ethics Board B. If you have any ethical concerns with the study, please contact Carleton University Research Ethics Board (by phone at 613-520-2600 ext. 4085 CUREB B or by email at [ethics@carleton.ca](mailto:ethics@carleton.ca)).

**Statement of consent – print and sign name**

I voluntarily agree to participate in this study.  Yes  No

\_\_\_\_\_  
Signature of participant

\_\_\_\_\_  
Date

**Research team member who interacted with the participant**

I have explained the study to the participant and answered any and all of their questions. The participant appeared to understand and agree. I offered to send a copy of the consent form to the participant for their reference.

\_\_\_\_\_  
Signature of researcher

\_\_\_\_\_  
Date