

Article

Shame Regulation in Learning: A Double-Edged Sword

Tanmay Sinha ^{1,*}, Fan Wang ² and Manu Kapur ^{2,*}¹ National Institute of Education, Nanyang Technological University, Singapore 637616, Singapore² D-GESS, ETH Zurich, 8092 Zurich, Switzerland; fan.wang@gess.ethz.ch

* Correspondence: tanmay.sinha@nie.edu.sg (T.S.); manukapur@ethz.ch (M.K.)

Abstract: Previous research and classroom practices have focused on dispelling shame, assuming that it negatively impacts self-efficacy and performance, and overlook the potential for shame to facilitate learning. To investigate this gap, we designed an intervention with 132 tertiary education students (45.46% male, 64.4% European ethnicity) spanning diverse undergraduate majors to show how and why designing for experiences of shame and appropriately regulating them can differentially impact learning. Shame was induced through autobiographical recall, imagination, and failure-driven problem-solving before randomly assigning students to three conditions: two with explicit tips for either decreasing shame or maintaining shame (*experimental groups*) and one with no-regulation tips (*control*). Students worked on an introductory data science problem deliberately designed to lead to failure before receiving canonical instruction. Manipulation checks triangulating self-reported and facial expression analysis data suggested that shame was successfully regulated in the intended direction, depending on the condition. Our results, drawing on mixed-methods analyses, further suggested that relative to students decreasing shame, those who maintained shame during initial problem-solving had (i) similar post-test performance on a non-isomorphic question and improved performance on the transfer question, evidenced by accuracy in solving applied data science and inference tasks; (ii) complete reasoning across all post-test questions, as evidenced by elaborations justifying the usage of graphical and numerical representations across those tasks; and (iii) use of superior emotion regulation strategies focused on deploying attention to the problem and reappraising its inherently challenging nature with an approach orientation, as evidenced by a higher frequency of such codes derived from self-reported qualitative data during the intervention. Decreasing shame was as effective as not engaging in explicit regulation. Our results suggest that teaching efforts should be channeled to facilitate experiencing emotions that are conducive to goals, whether they feel pleasurable or not, which may inevitably involve emoting both positive and negative (e.g., shame) in moderation. However, it is paramount that emotional experiences are not merely seen by educators as tools for improved content learning but as an essential part of holistic student development. We advocate for the deliberate design of learning experiences that support, rather than overshadow, students' emotional growth.



Academic Editor: Diego Vergara

Received: 9 February 2025

Revised: 15 April 2025

Accepted: 15 April 2025

Published: 17 April 2025

Citation: Sinha, T., Wang, F., & Kapur, M. (2025). Shame Regulation in Learning: A Double-Edged Sword. *Education Sciences*, 15(4), 502. <https://doi.org/10.3390/educsci15040502>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: emotion regulation; failure; shame; problem-solving

1. Introduction

Emotions inevitably color students' engagement with learning content, interactions with peers and teachers, appraisal of their performance, and reaction to feedback. Pekrun's seminal work on achievement emotions challenged the field to think about "what can be done to reduce students' (...) shame (...) and to foster their (...) enjoyment of learning?" (Pekrun, 2006, p. 337). Not surprisingly, empirical work on learning has focused primarily

on increasing positive emotions (e.g., [Belland et al., 2013](#); [Harley et al., 2017](#)). Ironically, although this literature acknowledges that negative activating emotions like shame may exert a variable effect on learning because of their potential to induce motivation to invest effort to avoid failure, the assumption that positive (pleasant) emotions should always be induced and prioritized because of their usefulness remains untested. To the best of our knowledge, no educational intervention has tested the causal effects of regulating shame in high-failure problem-solving situations. This mismatch between pedagogy and the assumptions regarding the usefulness of shame that drives its regulation motivates our work. We propose an intervention in data science to compare the relative efficacy of *downregulating* and *maintaining* shame on learning (RQ1—How does emotion regulation impact learning at the post-test?) and understand how students go through different stages of shame regulation (RQ2: How are different stages of emotion regulation distributed across the *downregulate*, *maintain*, and *no-regulate* conditions?).

In the remainder of this paper, we first describe our interdisciplinary theoretical background, linking two complementary bodies of work in the psychological science of emotions and contemporary learning science pedagogies that create opportunities for failure in the present intervention context. Next, we outline our study methodology, including participants, learning materials, design of our intervention conditions and measures, and our mixed-methods analysis plan comprising facial expression data for identifying emotion-related behaviors and quantitative coding of qualitative data, along with our hypotheses based on the frequency of these codes. In subsequent sections, we present our results and discussion, split along the lines of our two key research questions, focusing on *learning* and *shame regulation*. We showcase evidence for the implementation fidelity of our intervention and conclude with the implications of our work for educational theory and practice, limitations, and next steps in this exciting research on the intersection of cognition and emotion.

2. Theoretical Background

2.1. Shame Regulation

We situate our experimental manipulation of regulating shame within hedonic and counter-hedonic accounts of emotion regulation ([Tamir, 2016](#)). In simple terms, hedonic emotion regulation refers to strategies used to increase positive emotions or decrease negative emotions, while counter-hedonic emotion regulation refers to strategies used to decrease positive emotions or increase negative emotions. Prior research has relied primarily on hedonic theoretical accounts, ignoring the possibility that negative emotions can sometimes be fruitful in attaining instrumental task goals and that positive emotions can sometimes index complacency and be counter-productive in engaging meaningfully in the learning experience ([Sinha, 2022](#)). Counter-hedonic emotion regulation ([Tamir, 2016](#)), which better acknowledges the valence and usefulness of emotions, has largely not been explored in learning contexts. Generally, the degree to which emotions fluctuate over time plays an important role in well-being, and it is not always better to experience more positive emotions and fewer negative emotions ([Grant & Schwartz, 2011](#)), especially when the latter can serve as a reinforcer for learning and conditioning ([Baumeister et al., 2001](#)). More basic education research is therefore required.

During emotion regulation, people may “*increase, maintain, or decrease positive and negative emotions*” ([Koole, 2009](#), p. 6), which in turn affects their emotional responses, cognition, and behaviors. In classroom settings, reducing or avoiding students’ negative emotional reactions to learning content by framing it with positive emotional valence (e.g., via analogies and narratives ([Rosiek, 2003](#))) is a predominant approach. Teacher-student interactions are often directed at supporting students’ positive emotional experiences to

achieve classroom goals (Meyer & Turner, 2007), such as eliciting students' initial interest in engaging in learning activities. These approaches draw on traditional accounts of emotion regulation and focus primarily on scaffolding the hedonic need to promote pleasure and prevent pain as an intuitively plausible strategy (Harley et al., 2017; Kim & Pekrun, 2014; Quoidbach et al., 2015). However, when students' goal is to gain a deep understanding of challenging concepts, naturally occurring emotional struggles in problem-solving are catalysts for that learning goal. Two conflicting emotion regulation directions can be chosen in such scenarios: hedonic regulation to avoid the struggle and feel comfortable, or counter-hedonic regulation to remain uncomfortable and pave the way for deep learning. Theoretical frameworks of counter-hedonic emotion regulation emphasize that *"what we want to feel, therefore, and how we regulate our emotions may crucially depend on what we expect emotions to do for us"* (Tamir et al., 2015, p. 13). If teachers promote deep learning and resilience, arming students with knowledge of when emotions may be useful could increase their motivation to remain goal-oriented and engage in challenging but beneficial learning activities. These contrasting viewpoints motivated the design of our shame regulation prompts.

2.1.1. Shame Regulation Prompts in the Present Study

We created three experimental conditions that were aligned with the ethos of emotion regulation. Our first experimental condition focused on *downregulating* shame by using a two-pronged approach. First, we nudged the students to avoid feeling shame because other students in the past had shown this behavior. Second, we deployed students' attention away from shame by instructing them to focus on the positive emotions stemming from the reimbursement for study participation. Taken together, by putting a positive spin on task experiences, we hoped that students would re-interpret how they understood the emotional meaning of the situation, which would, in turn, decrease shame. This condition also helps explore the limits of focusing only on positive learning. It examines whether ignoring relevant, negatively valenced emotions tied to the learning experience, such as shame, can have harmful effects. Our second experimental condition challenged the first approach and focused on *maintaining* shame during the task by manipulating the participants' expected instrumental utility. Specifically, we framed shame as an emotion that students in the past found useful when they encountered task difficulties because it motivated them to reflect on and take action (or make amends) to restore their self-image. In situations where students are personally invested and have control over their goals, feeling ashamed of not reaching those goals can motivate them to work harder (Leach & Cidam, 2015; Turner & Schallert, 2001).

Across both the *downregulate* and *maintain* conditions, emotion regulation strategies were presented as tips and framed as case studies of similar peers, exemplifying an avoidance or approach orientation toward the event that caused shame. We posited that a social angle to framing the tip might intensify the effect of the regulation because of the likely weak (or non-) compliance with task suggestions perceived as more authoritative. Importantly, neither of these two conditions was intended to influence task progress through explicit performance expectations. Further, the tips in both conditions were presented via the common text *"You now have another opportunity to re-attempt the task and practice this tip. Please go ahead and revise your solutions"* (see Table 1), which sought to ultimately focus students' attention on the task, in one case by intentionally trying to make them experience shame (*maintain* condition), and in the other case by intentionally nudging them not to experience shame (*downregulate* condition). This common text was anticipated to counter any confounds stemming from participants' interpretation of the regulation tips as merely engaging in on-task (goal-oriented) versus off-task behaviors. Our final experimental

condition served as the control (*no-regulate*). Here, we did not use any regulation strategy and instead sought to better understand students' naturally occurring strategies.

Table 1. Emotion regulation prompts presented before the second problem-solving attempt.

Study Condition	Prompt
<i>Downregulate</i>	When other students in the past have attempted this task and encountered difficulties, they also felt shame. However, they did not dwell too much on it and instead felt happy that they were at least getting paid for the study. TIP: So, you may also try to avoid acknowledging that you may be feeling shame and always put a positive spin on your experiences. You now have another opportunity to re-attempt the task and practice this tip. Please go ahead and revise your solutions.
<i>Maintain</i>	When other students in the past have attempted this task and encountered difficulties, they also felt shame. However, they found shame to be useful because this led them to self-reflection and approach the task with even more focused efforts to restore their self-image. TIP: So, you may try to cultivate and experience shame in pursuit of your task goals and not assume the worst. You now have another opportunity to re-attempt the task and practice this tip. Please go ahead and revise your solutions.
<i>No-regulate</i>	You now have another opportunity to re-attempt the task. Please go ahead and revise your solutions.

2.1.2. Stages of Shame Regulation

We used the four stages of emotion regulation proposed by [Ford and Gross \(2019\)](#), building on [Gross \(2015\)](#), as the basis for understanding shame regulation using this framework. The *identification* stage serves as the starting point for individuals to detect and evaluate the emotions that require regulation. Several critical factors, such as (i) (in)adequate values attached to regulating an emotion (e.g., overvaluing excitement or pride, which may lead to difficulty in maintaining focus, potentially detracting from their effectiveness) or (ii) beliefs that emotions are immutable (difficult to change), can affect the fidelity of this stage. In the *maintain* condition, students are encouraged to embrace and utilize shame as a motivational tool for self-reflection and increased effort, as this tip reflects a form of emotion acceptance and utilization wherein students acknowledge and harness negative emotions for adaptive purposes; they may perceive it as less detrimental to their self-image. On the other hand, in the *downregulate* condition, students are encouraged to minimize acknowledgment of shame and adopt a positive reinterpretation of their task experiences (e.g., being compensated), which may lead them to perceive shame as an emotion that damages their self-image and needs to be avoided. Finally, in the absence of specific guidance on managing shame in a *no-regulate* condition, students defaulting to suppressing it ([Naragon-Gainey et al., 2017](#); [Tamir, 2016](#)) may implicitly stem from the common belief that shame damages their self-image.

The second stage of regulation is *selection*, where individuals evaluate the costs and benefits of an emotion regulation strategy, taking into account the available mental and physical resources, as well as the intensity of the emotion, and decide what strategy to use. Emotion regulation failures during this second stage may occur owing to both (i) a lack of consideration of contextual factors during strategy selection (e.g., short-term pleasure derived from avoiding challenging learning materials versus longer-term subject mastery) and (ii) poor self-efficacy beliefs to effectively use a particular strategy. During this stage, although we expect most students to find the tips' reasoning plausible (based on evidence from our pilot data), we also expect them to be uncertain about their helpfulness in problem-solving and/or find it difficult to use during their attempts (owing to relatively poor metacognition). We posit this to be difficult because students not only need to have (i) a nuanced contextualized belief about whether shame is (un)desirable that shapes their

behavior but also (ii) the appropriate conviction about the controllability of shame and (iii) the skill to apply the presented shame regulation strategy and continuously validate its impact on task performance (Ford & Gross, 2019; Gutentag et al., 2017). However, if some students initially perceive the regulation tips as not sensible, they may exhibit consistency in their negative attitude toward both the reasoning behind the tips and their practical problem-solving utility, as implementing emotion regulation strategies that they do not believe can be challenging.

The third stage is *implementation*, in which individuals consider various tactics for implementing emotion regulation strategies during a task (e.g., reframing setbacks as learning opportunities) by taking specific actions (e.g., seeing a low grade on a test as a chance to identify areas for improvement rather than as a reflection of incompetence). Challenges during the *implementation* stage may occur when individuals are unaware of or fail to perceive relevant tactics owing to a lack of practice, misjudging tactics in value, or simply struggling during implementation. Since students in the *maintain* condition are encouraged to embrace and accept their feelings of shame as part of the learning process and re-interpret it as a motivator for self-improvement, this may lead the majority of them to adopt a problem-solving orientation and confront the task directly with determination and resilience. In contrast, because students in the *downregulate* condition are encouraged to minimize their experience of shame and maintain a positive outlook, the majority may be more prone to interpreting their task performance in terms of its potential negative impact on their self-image and hence avoid approaching the challenging problem despite task instructions to engage, which aligns with the natural inclination to prioritize feeling good and avoid negative emotions (Naragon-Gainey et al., 2017).

The final regulation stage is *monitoring*, where individuals typically assess the progress and outcome of regulation, deciding “whether to maintain, switch, or stop ongoing regulation efforts” (Ford & Gross, 2019, p. 77). Beliefs about the goodness and controllability of emotions can affect regulation perseverance, which impacts the success of the *monitoring* stage. We posit that students in the *downregulate* condition will be more successful in their regulation attempts because of their alignment with the inherent human motivation to maximize pleasure and minimize pain (pro-hedonic motives (Naragon-Gainey et al., 2017; Tamir, 2016)). On the other hand, we expect students in the *maintain* condition to be relatively less successful in their regulation attempts owing to the rarity of the motivation for a decrease in pleasure or an increase in pain to promote hedonic balance (contra-hedonic motives (Naragon-Gainey et al., 2017; Tamir, 2016)).

Taken together, our interest in students’ qualitative appraisal of their emotion regulation experience led us to deploy a combination of deductive and inductive methods to develop a descriptive framework (Wang et al., 2024) based on the aforementioned emotion regulation stages (Ford & Gross, 2019). This descriptive framework allowed us to conjecture how shame regulation may proceed differentially across these stages for students in the *downregulate*, *maintain*, and *no-regulate* conditions. We outline these predictions explicitly in Section 3.6 and test them using our empirical analyses in Section 4.3.

2.2. Failure-Driven Learning Context

We used meta-analytic evidence (Sinha & Kapur, 2021b) on the effectiveness of *productive failure* (Kapur & Bielaczyc, 2012) to situate our study within this pedagogical framework. By engaging in problem-solving prior to instruction and exploring the problem-space for a yet-to-be-learned concept, students activate their prior knowledge, recognize deep problem features, become aware of knowledge gaps (Loibl et al., 2017). Despite encountering failure and negative affect during problem-solving (Lamnina & Chase, 2019; Sinha, 2022), students’ readiness to learn from follow-up instruction increases.

In our previous work, we proposed an optimized breakdown of the initial problem-solving phase within *productive failure* to enhance students' problem-space exploration (Sinha et al., 2021; Sinha & Kapur, 2021a). This involves a first sub-phase in which students freely generate solutions to a novel problem, followed by a second sub-phase in which students receive scaffolds that drive failure, which nudge the exploration of suboptimal representations related to the concept to be learned, leading to failed or suboptimal solutions by design. An example is working with one-dimensional histograms, which make it impossible to examine the covariation between variables, unlike a scatterplot. By complementing free generation with explicit scaffolding toward failure, students have been found to perform better in the post-test than in a prototypical single-phase design that only includes free generation (Sinha et al., 2021; Sinha & Kapur, 2021a). Taken together, scaffolds driving problem-solving failure can be conceived as challenge-oriented instructional support that may initially lead to failure but ultimately promote better learning.

Using culturally generalizable coding schemes for facial expressions (e.g., Cordaro et al., 2018), our secondary data analysis found that students who were scaffolded toward failure experienced shame, which was positively correlated with solution quality during initial problem-solving (Sinha, 2022). In ecologically valid learning situations, shame may affect the propensity of students to engage in solution and explanation generation during problem-solving, with the aim of achieving instrumental task goals (Leach & Cidam, 2015; Sinha, 2022). These goals focus on obtaining the desired outcomes, such as acquiring knowledge, rather than deriving pleasure from task completion. This contradicts the prevalent perspective that negative emotions should be avoided during learning and that increasing positive emotions is always preferable (Harley et al., 2017; Quoidbach et al., 2015). Recognizing that the scientific community holds diverse perspectives on the role of emotions in learning, our correlational findings on the potential benefits of shame call for the reconsideration of its role in learning and motivate further investigation into this understudied emotion. The causal influence of shame and its regulation on learning, which has not been fully understood in prior work, is the focus of the present study.

Shame Regulation and Learning

Productive failure pedagogical designs typically show evidence of learning via a post-test-only design that assesses outcomes like conceptual understanding and transfer (Kapur & Bielaczyc, 2012). Conceptual understanding outcomes assess whether students understand domain-specific ideas and why they work for certain problems, while transfer outcomes focus on whether and how students can apply domain-specific ideas to new but related problems that are not part of the initial training (Sinha & Kapur, 2021b). Post-test assessments of conceptual understanding can be broken down into items *isomorphic* to the initial problem-solving (e.g., same data and underlying problem-solving principle with a different cover story) and those that may be *non-isomorphic* to the initial problem-solving (e.g., the extension of the same underlying problem-solving principle with different data as well as different cover stories). See Sinha and Kapur (2021a) for more details.

When students solve post-test problems with lower intrinsic cognitive loads (e.g., conceptual understanding questions that are isomorphic to initial problem-solving), the motivation, additional engagement, and metacognitive awareness provided by experiences of shame may not be as helpful or necessary since the information required to solve the post-test problem is more readily apparent. However, continuing to experience shame during preparatory problem-solving tasks with a high intrinsic cognitive load may be an indication that students care deeply about the task or see it as an important part of their identity or values (Tracy & Robins, 2006). This may make such students (i) more willing to work through setbacks in order to achieve success and motivate them to persist through

the problem to (ii) understand the cause of their shame and develop new problem-solving strategies for the future, both of which may be instrumental in the success of higher-order transfer post-test tasks with similarly high levels of intrinsic load.

Irrespective of the post-test type, greater self-evaluation and reflection because of continuing to feel shame (Leach & Cidam, 2015) can lead to a deeper processing of information and drawing more connections across the generated representations. Students in the *maintain* condition can, therefore, be expected to show more complete reasoning in the post-test than those *downregulating* shame. Further, the differences between these two experimental conditions in the accuracy of their reasoning chain and the integration of graphical and numerical representations may also become more salient as the relative intrinsic load of the post-test questions increases (e.g., isomorphic versus non-isomorphic questions). Students *downregulating* shame may have fewer cognitive resources available to them due to reduced motivation or attention to divergent information coming from multiple representations (owing to the post-test design), which could, in turn, impair reasoning ability when solving post-test questions. Taken together, our choice of failure-driven learning design as the study context to situate our novel shame regulation intervention stemmed from its natural propensity to engage students in an emotionally colored, inquiry-driven learning experience (Sinha, 2022, 2025). Drawing on the longstanding history of research on failure-driven learning in the learning sciences (Sinha & Kapur, 2021b), this allowed us to make predictions on how our implemented shame regulation approaches may impact learning outcomes differently. The hypotheses are described in Section 3.6, and their support for our data sample is tested in Section 4.2.

3. Method

3.1. Participants

We recruited 132 participants from two European universities via email using a large pool of student volunteer information maintained by the university. This sample size was chosen based on apriori power analysis based on ANCOVA, which suggested $N = 128$ – 206 to detect medium-size effects (Cohen's d 0.5) with 70–90% power (α error probability 0.05). Each participant was compensated for 35 CHF for one and a half hours of participation time. Due to resource constraints, recruiting more participants was not feasible. Participants were required to have knowledge of high school math, basic computer skills, and consent to video recording. The sample comprised 45.46% males, 53.03% females, and 1.51% gender non-conforming or preferring not to answer. The majority of the participants (64.4%) were of European ethnicity, 31.06% were of Asian ethnicity, and 12.88% reported dual ethnicities. The participants had diverse undergraduate majors, with representation in the basic sciences (14.4%), applied sciences (24.24%, e.g., health science), engineering (22.7%), and humanities and social sciences (38.66%). This study was approved by the University Ethics Commission of *ETH Zurich* (EK-2022-N-192).

3.2. Learning Materials

We drew on previous work with validated learning materials on an introductory data science topic, *Anscombe's Quartet* (Sinha et al., 2021; Sinha & Kapur, 2021a). One learning goal was for students to understand the complementary importance of numerical and graphical representations when reasoning about data.

To achieve this goal, students received four bivariate datasets with variant-invariant features, a key feature of preparatory problem-solving within *productive failure* (Kapur & Bielaczyc, 2012). Information on the X-axis (units sold) and Y-axis (employee satisfaction) of the dataset, along with their similar non-parametric statistics, was provided as a starting point. Students were tasked with designing multiple measures to rank the datasets of the

four companies from most to least successful. A measure can take the form of a graphical (e.g., scatterplot) or numerical (e.g., computing averages) representation. These datasets had the same non-parametric statistics (e.g., median, interquartile range, Spearman's correlation) but possessed different parametric statistics (e.g., mean, standard deviation, Pearson's correlation) and contrasting visualizations. All participants made their first problem-solving attempt to generate multiple solutions for the task. As part of their second attempt at the problem, participants were sequentially presented with scaffolds that drive failure (one-dimensional histogram, bar graph, two-dimensional histogram; see (Sinha et al., 2021) and asked to engage with each scaffold to generate or revise measures. Across both attempts, the learning materials prompted the participants to be a safe space for idea generation and exploration. A web-based learning environment that allowed the use of predefined functions to create numerical representations and plot graphs was used (see Figure 1). An interactive tool tutorial was presented at the start for familiarization with all participants. See Supplementary Materials for further details.

INTRODUCTION <dataset description> <task description: design as many measures to rank order datasets>	ALLOCATED TIME <15 minutes>
TASK DATASETS <contrasting cases of multiple datasets with variant-invariant features>	KEY TERMS DEFINITION <e.g., median is...>
BASIC DESCRIPTIVE STATISTICS FOR TASK DATASETS <measures of central tendencies, spread, correlation>	PROCEDURAL HELP ACCESS <e.g., arithmetic/graphing functions>
	RECORDING SOLUTIONS <solution #1, reasoning, confidence> <solution #2, reasoning, confidence> <solution #3, reasoning, confidence> ...

Figure 1. Mock screen layout of the interactive problem-solving environment used in this study.

A previously validated *post-test* assessing conceptual understanding (one isomorphic and one non-isomorphic question) and transfer (one question), drawing on prior work (Sinha et al., 2021; Sinha & Kapur, 2021a), was used. The post-test evaluated the students' understanding of introductory data science concepts through varying complexities. The *isomorphic* question involved ranking three univariate datasets with identical boxplots based on their representation of socialist wealth distribution ideologies. The *non-isomorphic* question required students to identify a real dataset among trivariate datasets with the same regression fit but differing trends, emphasizing their understanding of real-world variable relationships. Finally, the *transfer* question challenged students to reconcile conflicting results from an inferential test metric (e.g., Shapiro-Wilk test), a descriptive metric (e.g., skewness), and visualizations (e.g., histogram) regarding air quality scores to reason about data normality. This post-test design implied that intrinsic cognitive load increased from *isomorphic* to *non-isomorphic* to *transfer* questions because of the increasing demand for participants' cognitive processes and their ability to apply knowledge in varied contexts. See the Supplementary Materials for details on these learning assessments.

3.3. Study Design

Our study utilized a *productive failure* learning design (see Figure 2) and started by gathering baseline data on participants' emotional intelligence, followed by an emotion induction procedure comprising autobiographical recall, imagination, and failure-driven problem-solving. Our procedure for *recalled shame* nudged participants to remember an experience within performance-focused contexts that had a high likelihood of evoking similar emotions as failure-driven problem-solving. Afterward, we framed the problem-solving task as an opportunity to engage in a new experience in which participants could restore their self-image and operationalize *imagined shame* by asking participants to imagine that their solution attempts would be peer-reviewed by another study participant who knew their recalled shameful event. Finally, students engaged in failure-driven problem-solving, which we conceive as *experienced shame*, drawing on our prior work that found observable evidence for the occurrence of this emotion within similar learning settings (Sinha, 2022). After the first attempt at the problem (15 min), a brief manipulation check was performed to assess the fidelity of the shame induction.

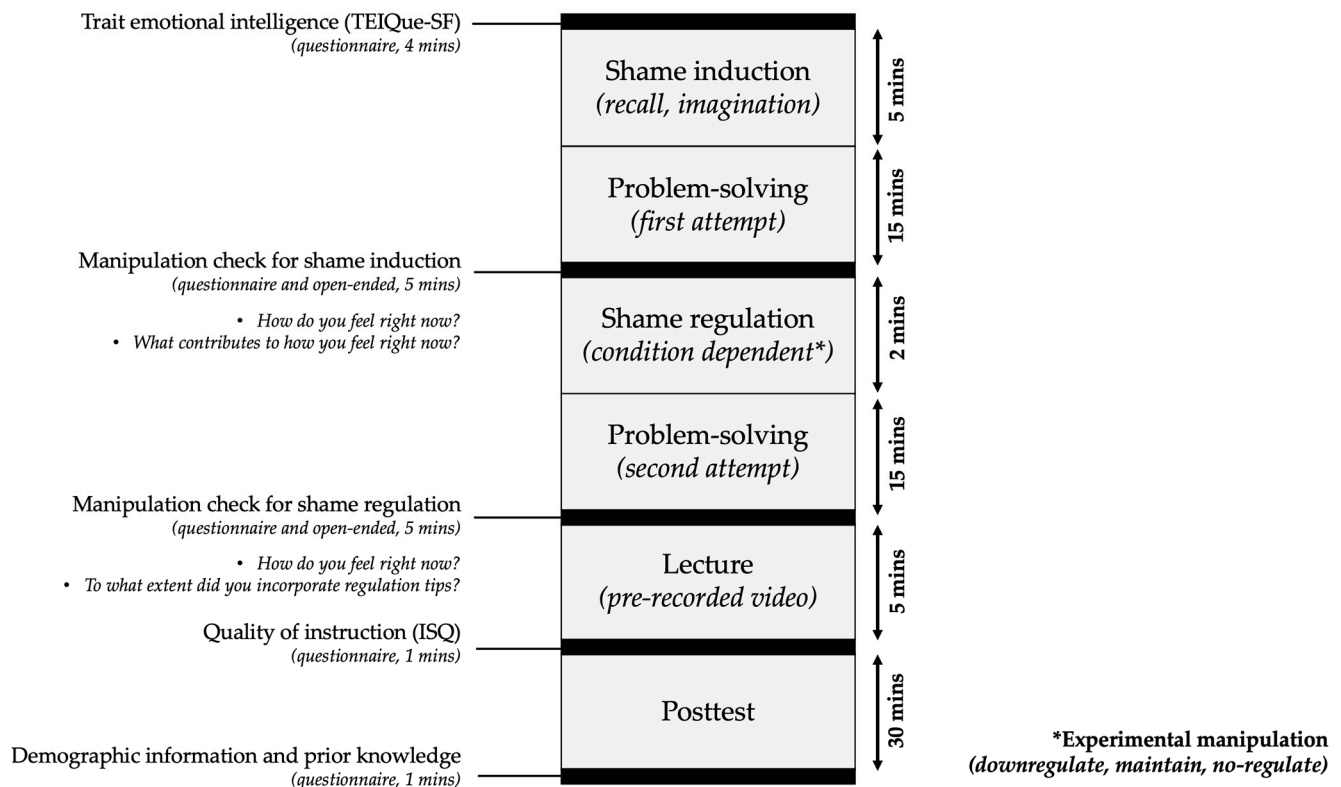


Figure 2. Study design showing the intervention phases and intermediary measurements.

Subsequently, the *emotion regulation* manipulation was instantiated where participants were randomly assigned to three experimental conditions before making a second attempt at the problem (15 min): *downregulate* ($N = 44$), *maintain* ($N = 44$), *no-regulate* ($N = 44$), and presented prompts, as shown in Table 1.

After the second problem-solving attempt (15 min), a manipulation check was administered to assess the fidelity of regulation, following which everyone received a common *lecture* on the targeted topic that contrasted common suboptimal ways of solving the problem with one canonical solution, aligning it with effective *productive failure* instruction (Sinha & Kapur, 2021b). Finally, the participants completed the post-test and reported their demographics, after which they were thanked for their participation and reimbursed for their time. Pilot testing with 12 participants was used to refine the learning materials,

procedure for induction and prompts for emotion regulation. In the online implementation of our study on Qualtrics, timers, and completion checklists were added to each section to keep the participants focused. See the Supplementary Materials for further details.

3.4. Measures

3.4.1. Before the First Attempt of the Problem-Solving Phase

We collected data on emotional intelligence using a previously validated trait questionnaire called the TEIQue-SF (Cooper & Petrides, 2010) within the first 4 min. TEIQue-SF assesses the perception of the ability to recognize, understand, and regulate emotions using a 30-item scale (5-point Likert scale ranging from *completely disagree* to *completely agree*). The scale demonstrated good internal consistency ($\alpha = 0.84$). Example items are “*Expressing my emotions with words is not a problem for me*” and “*I’m usually able to find ways to control my emotions when I want to*”. Subsequently, we asked participants to recall a prior university experience (e.g., problem-solving experience) in which they felt shame and to write down the event and the feelings they experienced (4 min).

3.4.2. Before the Second Attempt of the Problem-Solving Phase

After participants completed their first attempt at the task and subsequently re-read the descriptions of their recalled shameful event, our induction manipulation check comprised self-reports of whether participants felt ashamed, happy, satisfied, angry, full of contempt, disgusted, proud, bored, sad, or frightened at that moment (nine independent items, 5-point Likert scale ranging from *does not describe my feelings* to *clearly describes my feelings*). Participants were also asked to evaluate and explain the relative contributions of the following events to how they felt at that moment (scores summed to 100%): shameful event recall, imagining a peer knowing the shameful event, and participation in the problem-solving task.

3.4.3. After the Second Attempt of the Problem-Solving Phase

We asked participants to self-report whether they felt ashamed, happy, satisfied, angry, full of contempt, disgusted, proud, bored, sad, or frightened right now (nine independent items, 5-point Likert scale ranging from *does not describe my feelings* to *clearly describes my feelings*). In the *downregulate* and *maintain* conditions, participants were reminded of the tips about managing emotions they had received prior to the second attempt and asked (i) whether the tip seemed sensible, (ii) the extent to which they incorporated suggestions from the tip in their problem-solving (single item, 5-point Likert scale ranging from *I ignored the suggestions* to *I tried my best to incorporate*), and (iii) the rationale behind their approach, which served as the input for qualitative data coding to understand the differential impact of shame regulation and answer RQ2. To tap into naturally occurring emotion regulation strategies, participants in the no-regulation condition were simply asked (i) the extent to which they tried to manage their emotions when working through the problem (single item, 5-point Likert scale ranging from *I ignored my emotions* to *I tried my best to manage*), and (ii) the rationale behind their emotion regulation strategy.

3.4.4. After Instruction

After the participants completed the problem-solving phase and watched the instructional video, they rated the quality of the instruction using the shortened form of a previously validated questionnaire (ISQ; (Knol et al., 2016)) using a 7-item scale (5-point Likert scale ranging from *completely disagree* to *completely agree*). The scale demonstrated good internal consistency ($\alpha = 0.85$). This questionnaire tapped into facets like structure (e.g., “*the lecture was unorganized*”), explication (e.g., “*the instructor gave clarifying examples*”), and instruction (e.g., “*it was often unclear what the main and side issues were*”). Subsequently,

the participants completed the *post-test*. For each question, they could indicate answers and corresponding reasoning, which served as input for qualitative data coding to understand the different dimensions of post-test reasoning quality and answer RQ1.

3.4.5. After Post-Test

Finally, we asked participants to report demographic information, including gender, ethnicity, undergraduate major, familiarity with numerical and graphical representations (5-point Likert scale) as proxies for prior knowledge, and whether they had heard about the targeted concept of *Anscombe's Quartet* prior to the study (yes/no).

3.5. Analysis Plan

3.5.1. Facial Expression Data

To complement self-reports with rich multimodal process data on shame, we used facial expression analysis to probe deeper into the data and provide empirical evidence of the implementation fidelity of our intervention. Extending the multimodal learning analytics pipeline from [Sinha \(2022\)](#) and implementing it as an open-source tool in [Sinha and Dhandhanika \(2022\)](#), we performed a per-participant analysis to examine the incidence of facial expressions for shame and other co-occurring emotions. For example, in our work, based on [Cordaro et al. \(2018\)](#), the presence of an *international core pattern* of facial action units 4, 17, and 54 were considered indicative of an observable shame-like behavioral display comprising furrowed brows, raised lower lip, and head down. Such *international core patterns* serving as proxies for how emotions are expressed using facial movements have been shown to appear at above-chance rates across five cultures (China, India, Japan, Korea, and the United States), with no gender differences in their frequency of occurrence across these cultures. However, these prototypical emotional expressions only attempt to characterize behaviors that people readily associate with shame or other emotions when they see them in isolation rather than truly internal emotional states ([Cowen et al., 2021](#)). Indeed, alternative seminal perspectives on affective expression measurement address this issue ([Barrett et al., 2019](#)). In addition to examining emotion incidence, we constructed an emotional co-presence network to examine the relative importance of shame in facilitating the co-occurrence of other emotions by serving as a bridge or transition point. See the Supplementary Materials for more details.

3.5.2. Qualitative Data Coding

We coded two categories of open-ended responses to derive meaningful themes, as illustrated below and tested various explanatory hypotheses based on the frequency of these codes. We developed these coding schemes iteratively and evaluated the reliability of the codes with two raters with different ethnic backgrounds. First, we coded 20–30% of the data and resolved disagreements, after which one rater coded the remaining data. Coding was conducted blind to the students' ethnicity and condition.

Post-test reasoning quality—We expanded on prior work with similar learning materials ([Sinha & Kapur, 2021a](#)) to code reasoning quality along three dimensions (Table 2). First, *completeness* was evaluated, which included assessing whether students computed a graphical/numerical representation and elaborated on its mathematical meaning in isolation (lower levels of reasoning, codes P and E, Table 2, level 1) or whether they explicitly connected these descriptions to the problem-solving task and justified their usage (higher levels of reasoning, codes C and J, Table 2, level 1). Second, *accuracy* was considered, reflecting whether students were correct in their reasoning chain when engaging in the aforementioned higher levels of reasoning (Table 2, level 2). Finally, *integration* (Table 2, level 3) refers to students' ability to aggregate information from both numerical and graphical representations in their task reasoning.

Table 2. Coding the quality of post-test reasoning from verbal descriptions.

Level 1—Reasoning completeness			
Name	General definition	Label	Examples from the data
Presentation ($\kappa = 0.77$)	Participants presented or computed numerical or graphical representations.	P	Calculating the variance of a dataset, Drawing a scatterplot or histogram
Elaboration ($\kappa = 0.97$)	Participants described the mathematical meaning and trends in the data evident from drawing the numerical or graphical representations in isolation.	E	<i>“The index histogram is almost symmetric”</i> <i>“One dataset has larger variance in account balance which means it is not stable”</i>
Connection ($\kappa = 0.90$)	Participants explicitly connected the trends in the data with the problem-solving task.	C	<i>“Dataset B has rounded values for the TimeSinceFirstAccount attribute, which leads me to think that they were aggregated into bins to anonymise them.”</i> <i>“The skewness information represents that air quality index might be distributed symmetrically”</i>
Justification ($\kappa = 0.73$)	Participants supported the connection between their solutions and the problem-solving task with reasons for why the data trends can serve as plausible evidence.	J	<i>“It is unnatural for data points with both lower TimeSinceFirstAccount and lower AccountBalance to have darker color (higher CreditScore)”</i> <i>“The wealth owned by the community means that the middle class should be large, which is supported by the small variance of the wealth because the distribution is more concentrated”</i>
Level 2—Reasoning accuracy			
Name	General definition	Label	Examples from the data
Graphical accuracy ($\kappa = 0.95$)	Participants were accurate in their <i>connection</i> or <i>justification</i> for the graphical representations used/generated (cf. level 1).	GA	<i>“The graph distribution represents the assumption that the data is approximately distributed in a normal way”</i>
Numerical accuracy ($\kappa = 0.89$)	Participants were accurate in their <i>connection</i> or <i>justification</i> for the numerical representations used/generated (cf. level 1).	NA	<i>“If I understand the mayor’s ideology correctly, he wants wealth to be distributed as equally as possible. Since A has the lowest standard deviation, it should be the most representative scenario for the mayor’s ideology. According to the same criteria, scenario C is ranked second best and scenario B third best”</i>
Level 3—Reasoning integration			
Name	General definition	Label	Examples from the data
Graphical and numerical aggregation ($\kappa = 1$)	Participants combined both numerical and graphical representations that they used/generated in the final reasoning for the problem-solving task.	GN	<i>“The credit score is predicted by calculation the linear regression using the time since first account and the account balance. A higher account balance and longer time since opening an account should lead to a higher credit score, which is only clearly the case in dataset B. I chose to construct the graphs as I did because of the tutorial for adding a third variable and because of the regression equation, showing that the credit score is dependent on the other two variables.”</i> <i>“p-value of 0.03 is less than alpha of 0.05, which indicates rejection of H0, hence not a normal distribution. Skewness of air quality 0.0 indicates equal distribution in the right tail of the distribution and in the left tail of the distribution. By plotting the dots into bar groups, the graph shows a distribution very close to normal distribution. Combining the evidence above, it is ok to conclude that the air quality index scores across the cities are approximately normally distributed”</i>

Note. Double coding is possible across level 1 (e.g., P, E, C) and level 2 (e.g., GA, NA). Inter-rater reliabilities were averaged across post-test questions. Average scores across the first two levels were used as indices for *completeness* and *accuracy* in the statistical analyses.

Stages of shame regulation—Drawing on Ford and Gross (2019), we coded the four nuanced stages of shame regulation in different data passes. First, *identification* was assessed to determine whether students recognized their shame, as this awareness is a crucial

prerequisite for successful regulation (Table 3, level 1). Second, the *selection* was coded, examining how students, depending on their condition, considered whether and how to uptake the provided tip on shame regulation strategy (Table 3, level 2). Third, *implementation* was analyzed, describing how students approached shame regulation using concrete tactics during the problem-solving task (Table 3, level 3). Finally, *monitoring* was evaluated to assess the extent to which students attempted to evaluate the success or failure of their shame regulation within their metacognitive capacities (Table 3, level 4). Preliminary development of this coding scheme is available in Wang et al. (2024).

Table 3. Coding the stages of shame regulation.

Level 1—Identification			
Name	General definition	Label	Examples from the data
Self-image damage-related emotions ($\kappa = 0.67$)	Participants explicitly stated negative feelings about themselves during the problem-solving.	SD	"While doing the first attempt, I felt a bit helpless and stupid"
No self-image damage-related emotions ($\kappa = 0.80$)	Participants explicitly stated non-negative (neutral/positive) feelings about themselves during the problem-solving.	ND	"I did not feel a fair amount of shame"
Level 2—Selection			
Name	General definition	Label	Examples from the data
Agree with tip (tip_sensible = yes) ($\kappa_{\text{avg}} = 0.68$)	Participants acknowledged or agreed with the reasoning for the regulation strategy presented in the tip ($\kappa = 0.53$)	A1	"I think it is important to see things in a positive way even if they are not going so well" "Being reminded of how this is just a study, and not an exam that evaluates my abilities made me less stressed and happy that I'm getting money after this experiment"
	Participants found the information provided in the tip helpful for actually using it in their problem-solving task ($\kappa = 0.84$)	A2	"I felt better since I knew that I was a beginner with this tool and I needed guidance to kickoff from where I am" "Knowing that other people may feel the same made me feel better. I shouldn't think how the others are solving the problem in a better or worst way."
Disagree with tip (tip_sensible = no) ($\kappa_{\text{avg}} = 0.82$)	Participants did not acknowledge or agree with the reasoning for the regulation strategy presented in the tip ($\kappa = 0.84$)	D1	"How other people performed does not change the fact that I performed badly"
	Participants found the regulation strategy presented in the tip hard to achieve/use in their problem-solving task ($\kappa = 0.71$).	D2	"I found it hard to acknowledge the positive aspect in the fact that I'm getting paid for the study. The negative feeling due to the unclear exercise was too strong . . ." "I didn't quite understand how i should use my shame to achieve more satisfying results"
	Participants found the tip not applicable to them (e.g., because they did not feel shame) ($\kappa = 0.92$)	D3	". . .they also felt shame." But I actually don't. So it doesn't really apply to me. I am actually happy to. . ."

Table 3. Cont.

Level 3—Implementation			
Name	General definition	Label	Examples from the data
Attentional deployment as an emotion regulation strategy ($\kappa_{\text{avg}} = 0.78$)	Participants deployed their attention toward problem-solving (e.g., intentionally accepting the triggered emotions to focus on the task) ($\kappa = 0.80$)	AD1	“I gave myself time to complete the task without stressing out about it” “I just thought about doing my best without worrying too much about whether i am right or wrong” “I tried to be open-minded and think in new directions” “I do not know how to incorporate the tip into the assessment other than to remind myself that it is ok to be frustrated”
	Participants deployed their attention away from problem-solving (e.g., toward other positive aspects of the situation) ($\kappa = 0.76$)	AD2	“I acknowledge the positive aspect in the fact that I’m getting paid for the study”
Re-appraisal as an emotion regulation strategy ($\kappa_{\text{avg}} = 0.63$)	Participants tried to change the way they interpreted the problem-solving situation ($\kappa = 0.63$).	R1	“This is an open question without standard answers, so try without hesitation” “I struggled with knowledge rather than with execution or something else” “I focused on providing a better solution and perceived this as a challenge”
	Participants tried to change the way they interpret the influence of the problem-solving experience on their self-image ($\kappa = 0.63$).	R2	“Knowing that other people may feel the same made me feel better” “This also leads to a feeling of fulfillment even if the task did not turn out as though because I know I gave and did my best” “I don’t see the point being ashamed of something you have not done and tried before such as this case study”
Level 4—Monitoring			
Name	General definition	Label	Examples from the data
Ideal results	Participants explicitly reported a change in their emotions that was in accordance with their regulation efforts ($\kappa = 0.90$).	IR	“Knowing that other people may feel the same made me feel better” “Being reminded of how this is just a study, and not an exam that evaluates my abilities made me less stressed and happy that I’m getting money after this experiment.”
Non-ideal results	Participants explicitly reported attempting to regulate their emotions and did not achieve the intended outcomes ($\kappa = 0.48$).	NR	“The negative feeling due to the unclear exercise was too strong and I did not find the ideas very helpful either” “However it turns out I could not manage the emotions well”

Note. Double coding is possible across level 2 (e.g., A1 and A2) and level 3 (e.g., A1 and R1).

3.6. Hypotheses

We started out with a learning-focused question (RQ1) with theory-driven hypotheses about the effects of shame regulation on learning, drawing on Section Shame Regulation and Learning. Our interest in understanding students’ own appraisal of their shame regulation process (RQ2) further led us to collect textual reflection data, create coding schemes drawing on emotion regulation theory (see Section 2.1.2) to quantify the data, and develop post hoc hypotheses about the differential distribution of those codes in our sample (Chi, 1997).

RQ1 (*learning*): How does emotion regulation impact learning in the post-test?

- H1a (*accuracy*): With respect to the accuracy of the post-test solutions (percentage aligning with the canonical answer), participants in the *maintain* condition will score

(i) lower on the isomorphic conceptual understanding question, (ii) similar on the non-isomorphic conceptual understanding question, and (iii) higher on transfer questions, relative to the *downregulate* and *no-regulate* conditions.

- H1b (*reasoning quality*): With respect to the reasoning quality of the solutions developed at the post-test, participants in the *maintain* condition will score consistently higher in the completeness, accuracy, and integration dimensions of reasoning across all three post-test questions (relative to the *downregulate* and *no-regulate* condition).

RQ2 (*shame regulation*): How are different stages of emotion regulation (Table 3) distributed across the *downregulate*, *maintain*, and *no-regulate* conditions?

- H2a (*identification stage*): Fewer participants will identify shame and other unpleasant emotions as damaging their self-image in the *maintain* condition relative to the *downregulate* and *no-regulate* conditions ($SD_{downregulate}$ and $SD_{no-regulate} > SD_{maintain}$).
- H2b (*selection stage*): For participants who find the regulation tips sensible across all conditions, a relatively lower proportion of them will also acknowledge the helpfulness of the tip in their current problem-solving (A2) compared to those who agree with the reasoning presented in the tip (A1), that is, $A2 < A1$. Conversely, for participants across all conditions who do not find the regulation tips sensible in the first place, a similar proportion will disagree with the tips' reasoning and find them hard to use in their problem-solving attempts (D1–D2).
- H2c (*implementation stage*): Relative to the *downregulate* condition, a higher proportion of participants will deploy their attention toward problem-solving (AD1) by reinterpreting it with an approach orientation (R1) in the *maintain* condition. A reverse trend will hold with respect to the relatively maladaptive strategies of deploying attention away from problem-solving (AD2) and focusing too much on how problem-solving performance may damage participants' self-image (R2).
- H2d (*monitoring stage*): A higher proportion of participants in the *downregulate* condition would achieve their intended regulation results compared to those in the *maintain* condition ($IR_{downregulate} > IR_{maintain}$). Conversely, a higher proportion of participants in the *maintain* condition would be unable to achieve their intended regulation results relative to the *downregulate* condition ($NR_{maintain} > NR_{downregulate}$).

4. Results

4.1. Fidelity of Shame Regulation

Since our study focuses on the regulation of shame and its impact on learning, we briefly report our manipulation check, examining whether the change in the *ratio of self-reported shame to positive emotions* (happiness, satisfaction, and pride) before and after the second problem-solving attempt differed according to the experimental conditions. Because emotions are not completely independent and exist on a continuum, employing regulation strategies to increase the intensity of shame, for instance, may lead to a decrease in the intensity of positive emotions (and vice versa, see (Joseph et al., 2020)). The change in the *ratio of shame to positive emotions* reflects a metric that better accounts for this dual movement in negative/positive emotions resulting from emotion regulation. An ANCOVA analysis with averaged scores from the Trait Emotional Intelligence Scale as a covariate ($F(2, 128) = 2.731, p = 0.069, \eta_p^2 = 0.04$) suggested that the ratio of shame to positive emotions (i) decreased in the *downregulate* condition (marginal $M = -0.5, SE = 0.18$) as expected, even more so than the *no-regulate* condition (Cohen's $d = 0.28$); (ii) stayed fairly stable in the *maintain* condition (marginal $M = 0.09, SE = 0.18$) relative to the *downregulate* condition (Cohen's $d = 0.5$); and (iii) decreased only very slightly in the *no-regulate* condition (marginal $M = -0.17, SE = 0.18$). Considering only the changes in *shame* did not alter these trends.

When looking through the lens of a comparative facial expression analysis of attempts 1 and 2, 63.6% of the students ($n = 28$) in the *maintain* condition maintained or showed an increase in observable displays of shame-like behavior, while 86.36% ($n = 38$) and 75% ($n = 33$) of the students in the *downregulate* and *no-regulate* conditions maintained or showed a decrease in observable displays of shame-like behavior. A 10% threshold in the percentage of video frames tagged with shame-like behavior was used to assess the (dis)similarity of its incidence across attempts 1 and 2. The varying thresholds did not change the trends. Across our emotional network representation for the second problem-solving attempt drawn from facial expression analysis too, we found that the relative importance of facial displays of shame-like behavior, as assessed by the normalized *betweenness centrality* metric, was highest in the *maintain* (marginal $M = 0.031$, $SE = 0.005$) condition compared to the *downregulate* (marginal $M = 0.006$, $SE = 0.005$, Cohen's $d = 0.78$) and *no-regulate* (marginal $M = 0.0009$, $SE = 0.005$, Cohen's $d = 0.95$) conditions, even after controlling for the relative importance of the facial displays of shame-like behavior in the first attempt. Other manipulation checks corresponding to the fidelity of our learning task and shame induction can be found in the Supplementary Materials. In summary, these analyses showed that students indeed experienced failure in terms of not reaching the canonical answer and reported higher shame than other co-occurring emotions (including related emotions such as embarrassment). All measurements are listed in Table 4.

Table 4. Emotion measurements incorporated. The mean (standard deviation) is presented for self-reports and incidence. The normalized score is presented for the betweenness centrality metric.

	Self-Reports (Min 1, Max 5) (Questionnaires)	Incidence (Min 0%, Max 100%) (Facial Expressions)	Betweenness Centrality (Min 0, Max 1) (Facial Expressions Network)
First problem-solving attempt (before experimental manipulation)			
Shame	2.83 (1.38)	35.51 (28.37)	0.09
Happiness	1.98 (1.01)	1.35 (3.35)	0.01
Satisfaction	1.91 (1.04)	-	-
Anger	2.08 (1.18)	23.54 (24.33)	0.06
Contempt	2.06 (0.98)	12.40 (14.14)	0.03
Disgust	1.77 (1.02)	0.12 (0.55)	0.01
Pride	1.73 (0.89)	17.04 (23.05)	0.05
Boredom	2.46 (1.17)	-	-
Sadness	2.40 (1.29)	-	-
Fear	1.93 (1.11)	4.53 (7.88)	0.01
Amusement	-	1.35 (3.35)	0.02
Awe	-	1.18 (2.43)	0.02
Surprise	-	29.97 (28.56)	0.05
Confusion	-	23.54 (24.33)	0.06
Embarrassment	-	2.84 (7.67)	0.01
Pain	-	0.41 (1.16)	0.00

Table 4. Cont.

		Self-Reports (Min 1, Max 5) (Questionnaires)	Incidence (Min 0%, Max 100%) (Facial Expressions)	Betweenness Centrality (Min 0, Max 1) (Facial Expressions Network)
Second problem-solving attempt (after experimental manipulation)				
Shame	<i>Maintain</i>	2.70 (1.32)	30.05 (29.62)	0.03
	<i>Downregulate</i>	2.39 (1.40)	33.17 (32.14)	0.01
	<i>No-regulate</i>	2.89 (1.30)	29.15 (28.85)	0
Happiness	<i>Maintain</i>	1.98 (1.07)	1.69 (5.52)	0
	<i>Downregulate</i>	2.34 (1.03)	1.48 (4.86)	0.06
	<i>No-regulate</i>	2.11 (1.04)	1.65 (7.33)	0.04
Satisfaction	<i>Maintain</i>	2.02 (1.17)		
	<i>Downregulate</i>	2.41 (1.21)	-	-
	<i>No-regulate</i>	1.91 (1.14)		
Anger	<i>Maintain</i>	2.02 (1.30)	26.05 (28.97)	0.06
	<i>Downregulate</i>	1.50 (0.73)	22.10 (26.71)	0.04
	<i>No-regulate</i>	2.18 (1.33)	20.65 (23.48)	0.07
Contempt	<i>Maintain</i>	2.04 (0.96)	6.66 (9.90)	0.01
	<i>Downregulate</i>	1.89 (0.97)	9.83 (15.31)	0.13
	<i>No-regulate</i>	2.23 (1.18)	8.57 (16.03)	0.01
Disgust	<i>Maintain</i>	1.80 (1.10)	0.16 (0.89)	0
	<i>Downregulate</i>	1.41 (0.69)	0.14 (0.46)	0.01
	<i>No-regulate</i>	1.93 (1.13)	0.44 (2.78)	0
Pride	<i>Maintain</i>	1.75 (1.04)	20.36 (29.24)	0.08
	<i>Downregulate</i>	2.00 (1.01)	22.21 (29.48)	0.06
	<i>No-regulate</i>	1.80 (0.95)	14.21 (23.43)	0.07
Boredom	<i>Maintain</i>	2.54 (1.17)		
	<i>Downregulate</i>	2.45 (1.28)	-	-
	<i>No-regulate</i>	2.86 (1.27)		
Sadness	<i>Maintain</i>	2.07 (1.17)		
	<i>Downregulate</i>	1.68 (1.05)	-	-
	<i>No-regulate</i>	2.43 (1.30)		
Fear	<i>Maintain</i>	1.68 (1.03)	3.54 (6.93)	0.01
	<i>Downregulate</i>	1.39 (0.87)	6.09 (13.83)	0
	<i>No-regulate</i>	1.93 (1.06)	4.58 (10.04)	0
Amusement	<i>Maintain</i>		1.69 (5.52)	0
	<i>Downregulate</i>	-	1.48 (4.86)	0.07
	<i>No-regulate</i>		1.65 (7.33)	0.25
Awe	<i>Maintain</i>		1.23 (3.32)	0
	<i>Downregulate</i>	-	1.36 (3.33)	0
	<i>No-regulate</i>		1.09 (4.30)	0
Surprise	<i>Maintain</i>		30.43 (32.62)	0.15
	<i>Downregulate</i>	-	30.37 (34.01)	0
	<i>No-regulate</i>		36.35 (35.95)	0
Confusion	<i>Maintain</i>		26.05 (28.97)	0.06
	<i>Downregulate</i>	-	22.10 (26.71)	0.03
	<i>No-regulate</i>		20.65 (23.48)	0.07
Embarrassment	<i>Maintain</i>		3.35 (9.57)	0.01
	<i>Downregulate</i>	-	2.44 (6.33)	0.02
	<i>No-regulate</i>		2.37 (7.80)	0.03
Pain	<i>Maintain</i>		0.38 (0.97)	0
	<i>Downregulate</i>	-	0.45 (1.54)	0
	<i>No-regulate</i>		0.61 (2.98)	0.01

Note. Blanks (-) indicate that emotions were not measured using column indicators.

4.2. Learning (RQ1)

Our relatively small sample sizes led us to focus attention on Cohen's d measure of effect size to evaluate learning effectiveness. As an educationally relevant benchmark, Cohen's $d = 0.2$ can be considered practically important because it roughly equates to the yearly effect of having a teacher performing one standard deviation above the average (Hanushek, 2011). Results from ANCOVA analyses with post-test accuracy as the dependent variable and participants' familiarity with numerical and graphical representations as covariates showed the following:

- The *downregulate* and *no-regulate* conditions performed similarly (Cohen's $d = 0.10$) in the simplest isomorphic conceptual understanding question focusing on univariate data, while both these conditions outperformed the *maintain* condition (Cohen's $d = 0.45$ and 0.35 , respectively). The BF_c corresponding to this overall observed trend in marginal means (*maintain* < *downregulate* = *no-regulate*) was 14.36, indicating strong odds in favor of this informative hypothesis relative to its complement. In simple words, the Bayes factor of a specific descriptive hypothesis being tested versus its complement (BF_c) was obtained by running a *Bayesian informative hypotheses evaluation* ANCOVA (Hoijtink et al., 2019), which directly quantified the strength of evidence or odds in favor of the hypothesis.
- The relative differences between the three conditions drastically reduced for the non-isomorphic conceptual understanding question with relatively higher intrinsic cognitive load, with the *maintain* condition performing similarly relative to the *downregulate* (Cohen's $d = 0.11$) and *no-regulate* (Cohen's $d = 0.12$) conditions. The BF_c corresponding to this overall observed trend in marginal means (*maintain* = *downregulate* = *no-regulate*) was 53.32, indicating very strong odds in favor of this informative null hypothesis relative to its complement.
- Finally, as expected, trends completely reversed when looking at the transfer post-test question introducing novel ideas about hypothesis testing into reasoning with the problem and requiring the highest level of mental effort and processing—where the *maintain* condition outperformed both the *downregulate* (Cohen's $d = 0.13$) and *no-regulate* (Cohen's $d = 0.28$) conditions. The BF_c corresponding to this overall observed trend in marginal means (*maintain* > *downregulate* = *no-regulate*) was 10.99, indicating strong odds in favor of this informative hypothesis relative to its complement.

Taken together, these results support H1a, which posited lower, similar, and higher post-test accuracy for students in the *maintain* condition across the isomorphic, non-isomorphic, and transfer posttests relative to the *downregulate* and *no-regulate* conditions. Given the brevity and low-cost nature of our manipulation, the effects of higher-order transfer outcomes are worth highlighting because of their potential to accumulate over subjects and schooling years (Funder & Ozer, 2019), especially when a holistic emphasis is placed on regulation geared toward the maintenance and processing of negatively valenced emotions rather than downplaying them.

The results from ANCOVA analyses with the *completeness* dimension of *post-test reasoning quality* as the dependent variable and participants' familiarity with numerical and graphical representations as covariates showed the following:

- The *downregulate* and *no-regulate* conditions performed similarly (Cohen's $d = 0.04$) in the simplest isomorphic conceptual understanding question, but contrary to the trends in post-test scores, both these conditions were actually outperformed by the *maintain* condition (Cohen's $d = 0.25$ and 0.28 , respectively).

- The *maintain* condition scored similarly to the *no-regulate* condition (Cohen's $d = 0.10$) and further outperformed the *downregulate* condition (Cohen's $d = 0.18$) in the non-isomorphic conceptual understanding question.
- For the transfer question, the *maintain* condition was still superior to the *downregulate* condition (Cohen's $d = 0.33$) and the *no-regulate* condition (Cohen's $d = 0.32$).

The BF_c corresponding to the overall observed trend in the marginal means for the *completeness* dimension of *reasoning quality* across the three post-test questions was 14.80, 12.68, and 15.60, indicating strong odds aligning with informative hypotheses favoring the *maintain* condition relative to their complement. Again, these effects are up to $1.6\times$ stronger than educationally relevant benchmarks that draw on the average teacher's yearly impact (Hanushek, 2011). However, when looking at the remaining two reasoning quality dimensions, the only salient differences existed in the transfer question, in which participants from the *maintain* condition had the highest relative scores, outperforming the *downregulate* condition on both accuracy of reasoning (Cohen's $d = 0.23$) and integration of graphical and numerical representations in the solution (Cohen's $d = 0.18$). Taken together, these results partially support H1b, which posited higher scores for the *maintain* condition for all three dimensions of post-test reasoning quality (*completeness*, *accuracy*, and *integration*) across all post-test types, but only for the *completeness* dimension. All results, including the marginal means (standard errors), are summarized in Table 5.

Table 5. Marginal mean (standard error) of post-test scores and reasoning across conditions (RQ1).

	Isomorphic Conceptual Understanding	Non-Isomorphic Conceptual Understanding	Transfer
Post-test accuracy (percentage correct) (min 0, max 1)			
<i>Maintain</i>	0.62 (0.06)	0.46 (0.08)	0.66 (0.08)
<i>Downregulate</i>	0.81 (0.07)	0.51 (0.08)	0.59 (0.08)
<i>No-regulate</i>	0.77 (0.06)	0.52 (0.08)	0.52 (0.08)
Post-test reasoning quality (completeness) (min 0, max 1)			
<i>Maintain</i>	0.64 (0.04)	0.55 (0.04)	0.89 (0.08)
<i>Downregulate</i>	0.58 (0.04)	0.50 (0.04)	0.72 (0.08)
<i>No-regulate</i>	0.57 (0.04)	0.58 (0.04)	0.72 (0.08)
Post-test reasoning quality (accuracy) (min 0, max 1)			
<i>Maintain</i>	0.22 (0.04)	0.01 (0.02)	0.13 (0.04)
<i>Downregulate</i>	0.23 (0.04)	0.02 (0.02)	0.07 (0.04)
<i>No-regulate</i>	0.21 (0.04)	0.02 (0.02)	0.10 (0.04)
Post-test reasoning quality (integration) (min 0, max 1)			
<i>Maintain</i>	0.02 (0.02)	0.02 (0.03)	0.43 (0.07)
<i>Downregulate</i>	0.02 (0.02)	0.02 (0.03)	0.34 (0.07)
<i>No-regulate</i>	0.02 (0.02)	0.05 (0.03)	0.36 (0.07)

In addition to these learning results, we also found in post hoc exploratory analyses that after controlling for students' prior knowledge (familiarity with graphical and numerical representations) and the extent to which they incorporated the emotion regulation tips in their problem-solving, students in the *maintain* condition grappled with the presented lecture for a relatively longer time (Cohen's $d = 0.23$) and appraised it as having a higher instructional quality than students in the *downregulate* condition (Cohen's $d = 0.28$).

4.3. Shame Regulation (RQ2)

First, with regards to the *identification* stage of regulation, evidence from a one-tailed independent samples *t*-test ($t(130) = 1.713$, $p = 0.045$, Cohen's $d = 0.32$) suggested a lower proportion of codes signaling explicit verbalization of self-image damage-related emotions in the *maintain* condition ($M = 0.23$, $SD = 0.42$) compared to the *downregulate* and *no-regulate* conditions taken together ($M = 0.37$, $SD = 0.49$). This supports H2a ($SD_{downregulate}$ and

$SD_{no-regulate} > SD_{maintain}$), which expected fewer participants in the *maintain* condition to perceive shame and other unpleasant emotions as being harmful to their self-image.

Second, with regards to the *selection* stage of regulation for those who found the presented tips sensible, evidence from a one-tailed one-sample *t*-test ($t(39) = -5.234$, $p < 0.001$, Cohen's $d = -0.83$) suggested a sharp drop in codes that also acknowledges the helpfulness of tip during immediate problem-solving ($M = 0.22$, $SD = 0.42$) compared to codes that simply agreed with the reasoning of the tip ($M = 0.57$, $SD = 0.50$). For those who did not find the tip sensible, although, a similar statistical analysis ($t(47) = 0.698$, $p = 0.489$, Cohen's $d = 0.10$, $BF_{01} = 5.07$) showed no difference in the proportion of codes belonging to disagreeing with the tips' reasoning ($M = 0.21$, $SD = 0.41$) and finding it hard to use during immediate problem-solving ($M = 0.17$, $SD = 0.38$). In fact, the posited null hypothesis was 5.07 times more likely to be true than the alternative hypothesis. Taken together, these results support H2b ($A2 < A1$ and $D1-D2$), which posited that participants' acknowledgment of the regulation tips' helpfulness in problem-solving would be lower than or similar to merely agreeing with the tips' reasoning, depending on the initial appraisal of the tip as sensible.

Third, with regards to the *implementation* stage of regulation, evidence from a one-tailed independent samples *t*-test ($t(86) = 1.349$, $p = 0.090$, Cohen's $d = 0.29$) suggested that participants in the *maintain* condition ($M = 0.25$, $SD = 0.44$) had a higher proportion of verbalizations comprising the regulation strategy of deploying attention toward the problem than those in the *downregulate* condition ($M = 0.14$, $SD = 0.35$). A similar trend was observed for the emotion regulation strategy of reappraising the inherently challenging problem-solving situation as a way to continue making progress ($t(86) = 1.617$, $p = 0.055$, Cohen's $d = 0.34$), with participants in the *maintain* condition ($M = 0.18$, $SD = 0.39$) having a higher proportion of such verbalizations relative to the *downregulate* condition ($M = 0.07$, $SD = 0.25$). When analyzing evidence for the remaining two relatively maladaptive regulation strategies, participants in the *maintain* condition further showed not only a (i) lower proportion of verbalizations comprising attentional deployment away from the problem ($M = 0.02$, $SD = 0.15$) relative to the *downregulate* condition ($M = 0.25$, $SD = 0.44$), as evidenced by a one-tailed independent samples *t*-test ($t(86) = -3.254$, $p < 0.001$, Cohen's $d = -0.69$), but also a (ii) lower proportion of verbalizations comprising re-appraisal focused primarily on self-image perception and repair ($M = 0.14$, $SD = 0.35$) relative to the *downregulate* condition ($M = 0.25$, $SD = 0.44$), as evidenced by a one-tailed independent samples *t*-test ($t(86) = -1.349$, $p = 0.090$, Cohen's $d = -0.29$). Taken together, these results support H2c ($AD1_{maintain} > AD1_{downregulate}$ and $R1_{maintain} > R1_{downregulate}$ and $AD2_{downregulate} > AD2_{maintain}$ and $R2_{downregulate} > R2_{maintain}$), which posits a relatively greater focus on problem-solving by adopting an approach-oriented perspective for students in the *maintain* condition.

Finally, with regards to the *monitoring* stage of regulation, evidence from a one-tailed independent samples *t*-test ($t(86) = 1.239$, $p = 0.109$, Cohen's $d = 0.26$, $BF_{01} = 1.31$) showed a similar proportion of codes where participants achieved their intended regulation results across the *maintain* ($M = 0.09$, $SD = 0.29$) and *downregulate* ($M = 0.18$, $SD = 0.39$) conditions. Although descriptively aligning with H4d, the odds of the null hypothesis were only 1.31 times higher than those of the posited alternative hypothesis. An analogous analysis of the code signifying the inability to achieve the intended emotion regulation results ($t(86) = 1.418$, $p = 0.920$, Cohen's $d = 0.30$, $BF_{01} = 9.96$) showed a similar proportion across the *maintain* ($M = 0.11$, $SD = 0.32$) and *downregulate* ($M = 0.23$, $SD = 0.42$) conditions, with strong odds of 9.96 favoring the null hypothesis as opposed to the posited alternative hypothesis. Taken together, these results do not support H2d, which predicts a differential distribution of monitoring codes across conditions ($IR_{downregulate} > IR_{maintain}$ and $NR_{maintain} > NR_{downregulate}$).

5. Discussion

We implemented a high-fidelity intervention ($N = 132$) to understand how the differential regulation of shame via *downregulation* and a contrasting *maintenance* approach affects learning. We utilized a failure-driven approach and self-reports, along with facial expression analyses, in our interdisciplinary education research.

5.1. Learning (RQ1)

Our results suggest that although *maintaining* shame may not serve students well for simpler (isomorphic) conceptual understanding problems, *maintenance* may not be relatively worse (does not hurt) for performance on non-isomorphic conceptual understanding problems and may, in fact, be better than *downregulating* shame for transfer problems requiring sustained effort and attention, and a more thorough understanding of the learning materials. This is an important finding because it suggests that counter-hedonic shame regulation may play a different role in learning depending on the complexity of the problem. Continuing to experience shame, as was the focus of our regulation manipulation in the *maintain* condition, may also play an important role in making students more likely to identify gaps or inconsistencies in their thinking, leading to more complete reasoning, as was evident in the data from all three post-test questions. Taken together, counter-hedonic shame regulation may facilitate paying greater attention to the details of the task and thinking more critically about one's reasoning, resulting in explicitly verbalized connections between the generated representations and the task and observable effort toward solution justification. Continuing to experience (as opposed to downregulating) shame during problem-solving may also serve another collateral benefit, as evidenced in our post hoc exploratory analyses of how students in the *maintain* condition appraised the lecture and suggested increased receptivity to learning from it. We posit that this may be a function of students' enhanced (i) focus on evaluating the canonical solution and (ii) openness to receiving guidance, both of which may contribute to deeper information processing (e.g., see (Naismith & Lajoie, 2018) for a similar design where shame predicted attention to expert feedback).

Optimism plays an important role in increasing perseverance in terms of the occurrence and frequency of study behaviors (Seligman, 1991) and correlates with better psychological well-being by buffering stress (Scheier & Carver, 1985). Although cultivating such a positive emotional framing could sometimes be a powerful coping mechanism, putting a positive spin on a challenging problem-solving situation and discrediting the negative may also be counter-productive. Because it encouraged an avoidance orientation toward the event that caused shame en route to reworking the second problem-solving attempt, the *downregulation* condition may have implicitly nudged students toward toxic positivity (Ford & Mauss, 2014). The concept of toxic positivity refers to the idea that one should always maintain a positive outlook and avoid negative emotions, even in challenging situations. Prioritizing feeling good over actively engaging with the discomfort associated with learning something new, after all, aligns with the natural human inclination (Naragon-Gainey et al., 2017). However, this may be problematic as it can lead to a lack of acknowledgment of negative emotions and experiences and habitually judging them as bad/inapt (Willroth et al., 2023), which can ultimately hinder personal growth and learning. The *maintain* shame condition, conversely, not only validated students' emotional experience but also made them aware of the utility of experiencing it, thereby intentionally nudging students to try and feel shame en route to reworking on the second problem-solving attempt. Prior work has shown that knowledge of when particular emotions are useful may lead people to prefer and cultivate emotions that aid in goal attainment, irrespective of whether they are pleasant or unpleasant (Tamir et al., 2015).

Interestingly, we had not hypothesized apriori that our learning results would suggest similar performance for students across the *no-regulate* and *downregulate* conditions. This was evidenced in the accuracy scores for all three post-test questions and reasoning completeness for two of the three questions. What this counter-intuitively suggests is that sometimes (e.g., in cases where implementation of shame maintenance may not be feasible), it may be better not to intentionally design for an immediate reduction in the intensity or frequency of students' shame experiences but *leave it brewing* (at least initially!), as this does not seem to be empirically disadvantageous for learning. Now, while doing so may give students agency to handle shame on their own without creating emotional dependency on comfort from the external world, it inevitably runs the risk of students reverting back to their natural tendencies of decreasing negatively valenced emotions (Naragon-Gainey et al., 2017; Tamir, 2016). However, we note that this decrease may be relatively quite mild, as evidenced by our manipulation checks suggesting a larger decrease in the ratio of shame to positive emotions for the *downregulate* condition compared to the *no-regulate* condition (Cohen's d 0.28) after the introduction of tips.

5.2. Shame Regulation (RQ2)

Our coding results on four stages of shame regulation suggested that during the *identification* stage, students in the *maintain* condition had few verbalizations signaling damage to their self-image. This is plausible because such students are likely to see unpleasurable emotions such as shame in a more positive light (owing to the presented tip) and, therefore, will not necessarily identify shame as damaging self-image (e.g., "*The tip helped me realize that it is ok to make mistakes, that what matters is the effort, and the fear of being ashamed because of a mistake can be used as a motivation*").

During the *selection* stage, we found that for the $n = 40$ students who found the presented emotion regulation tips sensible across the *downregulate* and *maintain* conditions, only one-fourth ($n = 9$) also acknowledged the tip to be actually helpful during their problem-solving attempt (e.g., "*With the tip I know I'm not the only one who feels confused during the problem solving process, which makes me feel less ashamed when I come across problems*"). Students need to have good metacognitive awareness to be able to assess the helpfulness of applying emotion regulation tips in the task, which may be scarcely prevalent in university students (e.g., "*I tried to engage this ashamed feeling into the problem solving process. But it is a bit difficult to always push myself to think about shame. The difficulty is to believe this feeling of shame is actually true. . .*"). Designing interventions to address this issue is important for future work.

With regards to the *implementation* stage, we found that a greater proportion of students in the *maintain* condition used relatively superior emotion regulation strategies such as deploying attention toward the problem (e.g., "*I decided to try to think even better about the problem by trying to focus more on the topic and in what was asked*") and positively reappraising the challenge inherent in the task (e.g., "*Well being overwhelmed won't help me so I somehow manage to accept the situation and just try my best to understand and find out some new things in this program. I might not be able to give 'good' answers but I can at least have a little fun with the program. . .*"). This may have contributed to increased focus in trying to generate new solutions without fear of embarrassment or failure (e.g., "*I wanted to have some positive reinforcement from feeling dissatisfied and not being overly confused. Therefore, I thought about the ideas and used them. I was a bit happy that I was able to perform the first idea even before the idea tutorial*"), which in turn made these students more receptive to future instruction and led to similar or better post-test performance than students in the *downregulate* condition. On the contrary, focusing attention away from the task or thinking too much about how the problem-solving performance may impact their self-image (e.g., "*Earlier I was taking too long*")

to find any solutions and taking a lot of time to think. I later realized that if I take it less to heart then it's better because I will still get paid the same, so I should stay calm and not overthink this", "I tried to avoid acknowledging that I feel ashamed but that didn't really help me in getting the tasks done. I still didn't know how to solve the problems, I just took it with a grain of salt and didn't get too involved with it anymore"); students in the *downregulate* condition could have been distracted, which in turn prevented them from exploring the problem-space fully, thus reflecting in poor post-test performance on the questions with higher intrinsic cognitive load.

Finally, during the *monitoring* stage of emotion regulation, we did not find verbalized evidence of a greater proportion of students in the *downregulate* condition achieving their intended regulation results relative to the *maintain* condition. Our rationale for this hypothesis was rooted in prior empirical evidence suggesting that downregulating negative emotions is commonplace, but maintaining or upregulating them is rare, harder to implement, and presumably counter-intuitive (Naragon-Gainey et al., 2017; Tamir, 2016). However, a qualitative examination of responses suggested that the relatively stronger signal from problem-solving may have introduced more discrepancies for students in the *downregulate* condition. Despite their initial efforts to downregulate negative emotions, challenges encountered during the problem-solving process could have amplified the inconsistencies between their nudged emotional state and the difficulties these students actually faced, leading to non-ideal monitoring outcomes (e.g., *"It did help a bit to remember that I'm getting paid. However, I feel even more ashamed now because I feel like I failed even with the help on my 2nd attempt(s) so the negative feelings definitely outweigh the positive. It's nice that I'm getting paid but I still feel very stupid"*). This trade-off may have, in turn, dampened the routineness of downregulating shame.

5.3. Implications for Theory and Practice

We push research on preparatory approaches for future learning (Kapur & Bielaczyc, 2012; Loibl et al., 2017; Sinha & Kapur, 2021b) forward by investigating understudied affective dimensions that may strongly contribute to students' learning. Showcasing the benefits of counter-hedonic shame regulation on learning, our findings implicate and invite the development of theory for teasing apart the valence of emotions from their usefulness within ecologically valid learning contexts. Valence, a primitive emotional dimension, refers to the inherent positive (e.g., happiness, joy, pride) or negative orientation (e.g., shame, anger, confusion) of the emotion. Usefulness, on the other hand, refers to the instrumental quality of an emotion, in particular, its relevance in accomplishing a specific goal in a given context (Tamir, 2016). For example, a negatively valenced emotion like anger is likely to be most useful in negotiating with a stranger, when engaging with problem-solving in the context of social injustice, or when attaining challenging learning goals (Holbert, 2020; Lench et al., 2024; Tamir et al., 2015). Similarly, positively valenced emotions like joy and pride are less likely to be useful when an individual wants to appear humble during victory and prioritize empathy with relatively poor-performing peers over viewing themselves favorably (Kalokerinos et al., 2014). Similar empirical evidence is beginning to be echoed in other research on cognition and affect in areas like stress (e.g., Travis et al., 2020) and passion (e.g., Vallerand et al., 2023).

A direct implication for educators, then, is how to scaffold students' usage of *emotions as tools* (Weidman & Kross, 2021), that is, nudge them to try and feel/express a certain emotion, not because of its positive or negative valence, but because of the strong match between its functional purpose and the situational demand. In contrast to subjective well-being research (Diener, 2000) that posits the more positive emotions we experience, the happier we are, this view implicates designing for learning involving experiences of emotions that are goal-conducive, whether they feel good or not. Constantly seeking and

overvaluing happiness may, after all, sometimes have paradoxical effect of making people unhappy (Mauss et al., 2011).

For our work, this begs the key question of better utilization of classroom resources to facilitate learning from a broader repertoire of instrumentally useful emotions like shame. Recall that our goal is not to intentionally (and unethically!) shame students for underperformance but rather to offer a diverse educator toolbox that incorporates reflections of shame into learning in moderation and in the right context by drawing from students' own experiences as a starting point. With emotion regulation guidance focused on goal-oriented re-appraisal in challenging failure-driven learning contexts, the process of constructively approaching shame may actually benefit learning more in contrast to simply shedding it off. In fact, discounting shame can deter students from seeking social support (Tangney & Tracy, 2012). Therefore, our empirical evidence can support teachers' and students' choices of when and what emotions to condemn or cultivate to attain task goals. While feeling shame may not be the ultimate long-term goal, experiencing it may play a significant role in catalyzing the necessary changes (e.g., better time management, encouragement of accountability, and enhanced empathy) to achieve task goals.

We propose that teachers help their students grow through emotional experiences by changing their perspectives on negatively valenced emotions. Specifically, teachers can (i) encourage sharing and building support systems (e.g., *"what's something that made you upset or shameful when you were trying to learn? Let's talk about it together. I'm here to listen and help you work through them"*), (ii) validating and normalizing feelings (*"It sounds like you really felt disappointed when that happened. I can see why you felt that way—learning can be tough. It's hard when things don't go as we hope"*, or, *"It's completely normal to feel frustrated or embarrassed when learning something new. It shows that you care about doing well. Everyone goes through that at some point"*), and (iii) reframing feelings and collaboratively setting goals (*"Instead of letting that shame hold you back, how can we use that feeling to motivate you to try again? Let's set a small goal to work on. What's one thing you'd like to improve?"*, or, *"What if we think of that frustration as a sign that you're pushing your limits? That's actually a good thing. Maybe we could create a plan to turn this negative feeling into a positive action?"*). Creating space for such empathetic conversations can reorient students' focus and encourage them to harness negatively valenced emotions in a way that helps them attain task goals.

5.4. Limitations and Future Work

Our short intervention examined only the immediate effects of shame on learning, leaving much to be desired in terms of understanding the developmental trajectory, carryover, and habituation effects of shame on achievement. Despite our proactive attempt to prevent confounds arising from goal orientation in the framing of the regulation tips, it still remains a follow-up study to alleviate this concern in future work—here, one may add a comparison condition that more explicitly nudges students to approach the problem while focusing on experiencing pleasurable emotions instead of shame. A post-tip measure assessing willingness to re-attempt the problem (e.g., on a 5-point Likert scale) can serve as a manipulation check.

Our study design does not allow making direct inferences on the occurrence of shame in *productive failure* designs when participants are not being prompted by explicit induction procedures (e.g., *recalled shame*, *imagined shame*). However, when taken together with existing correlational research based on similar learning materials and demographics, and that included such a baseline condition where students naturally experienced a wide palette of emotions, including shame (Sinha, 2022), our study trajectory began to build strong empirical evidence for the causal role of emotions in *productive failure* learning designs. Finally, although we attempted to tap into shame using verbal self-reports and

facial expressions, there is a need for more nuanced measures to capture shame-related constructs (e.g., withdrawal tendencies) and emotional regulation strategies in ecologically valid learning contexts (e.g., [Sinha, 2025](#)). Methodologically, our small sample sizes may have reduced the statistical power to detect true effects, leading to biased estimates of effect size; however, we attempted in our analysis plan to not solely rely on null hypothesis significance testing but also incorporate Bayes factors to transparently provide the strength of evidence favoring various null/alternative hypotheses across our research questions. Replication with more students across various cultural geographies would strengthen the generalizability of our findings.

What may be the focus of future research into emotion regulation in learning? One area worth considering is how to scaffold sensemaking with useful emotions to improve the effectiveness of constructivist designs, such as *productive failure*. For instance, can case studies and narratives from peers be used to offer vicarious learning opportunities where students engage in contrasting approach/avoidance behaviors toward similar academically relevant emotions? How can established psychological methods like expressive writing—where students engage in structured reflections about navigating emotions during challenging learning tasks—be used to allow students to actively generate the utility of useful emotions in learning? Can we tap on the rich collaborative learning literature to engage students in argumentation with peers in low-stakes settings so that they can co-construct meaning from negatively valenced emotions like shame? We hope that the present investigation will spur scientific inquiry along these lines.

6. Conclusions

We challenged normative perceptions of shame in learning and explored emotion regulation strategies to downregulate or maintain shame. Our results suggest the need for further research on the complex interplay between counter-hedonic emotion regulation and performance among students. We call for the development of interventions and the investigation of other emotions that may naturally be present in challenging learning contexts but are normally perceived as unpleasurable (e.g., anger, disgust, and contempt) and typically dispelled by educators to maintain a positive climate in the classroom. Our work highlights the potential benefits of facilitating goal-conducive emotions, whether positive or negative, in moderation.

Supplementary Materials: Supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/educsci15040502/s1>.

Author Contributions: Conceptualization, T.S.; methodology, T.S. and F.W.; validation, T.S.; formal analysis, T.S.; investigation, T.S. and F.W.; resources, T.S. and M.K.; data curation, T.S. and F.W.; writing—original draft, T.S.; writing—review & editing, T.S., F.W. and M.K.; supervision, T.S.; project administration, T.S. and M.K. All authors have read and agreed to the published version of the manuscript.

Funding: T.S. is supported by the National Institute of Education, Singapore under a Start-Up grant (NIE-SUG 5-23 TS). The APC was funded by ETH Zurich.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of ETH Zurich EK (protocol code 2022-N-192 and 11 October 2022).

Informed Consent Statement: Informed consent was obtained from all the subjects involved in the study.

Data Availability Statement: All anonymized data, except for sensitive facial recording data, are available upon reasonable request from the corresponding authors.

Acknowledgments: We thank Stefan Wehrli, Peruntha Mohanarasa and Cornelia Schnyder from the Decision Sciences Lab (ETH Zurich) for their assistance in running the study, automating the study workflow, and handling participant recruitment and reimbursement. Thanks to Maya Spannagel and Sofia Strukova for assistance with coding the data.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, 20(1), 1–68. [\[CrossRef\]](#)
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4), 323–370. [\[CrossRef\]](#)
- Belland, B. R., Kim, C., & Hannafin, M. J. (2013). A framework for designing scaffolds that improve motivation and cognition. *Educational Psychologist*, 48(4), 243–270. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *Journal of the Learning Sciences*, 6(3), 271–315. [\[CrossRef\]](#)
- Cooper, A., & Petrides, K. V. (2010). A psychometric analysis of the Trait Emotional Intelligence Questionnaire–Short Form (TEIQue–SF) using item response theory. *Journal of Personality Assessment*, 92(5), 449–457. [\[CrossRef\]](#)
- Cordaro, D. T., Sun, R., Keltner, D., Kamble, S., Huddar, N., & McNeil, G. (2018). Universals and cultural variations in 22 emotional expressions across five cultures. *Emotion*, 18(1), 75–93. [\[CrossRef\]](#)
- Cowen, A. S., Keltner, D., Schroff, F., Jou, B., Adam, H., & Prasad, G. (2021). Sixteen facial expressions occur in similar contexts worldwide. *Nature*, 589(7841), 251–257. [\[CrossRef\]](#)
- Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. *American Psychologist*, 55(1), 34–43. [\[CrossRef\]](#)
- Ford, B. Q., & Gross, J. J. (2019). Why beliefs about emotion matter: An emotion-regulation perspective. *Current Directions in Psychological Science*, 28(1), 74–81. [\[CrossRef\]](#)
- Ford, B. Q., & Mauss, I. B. (2014). The paradoxical effects of pursuing positive emotion: When and why wanting to feel happy backfires. In J. Gruber, & J. T. Moskowitz (Eds.), *Positive emotion: Integrating the light sides and dark sides* (pp. 363–381). Oxford University Press. [\[CrossRef\]](#)
- Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, 2(2), 156–168. [\[CrossRef\]](#)
- Grant, A. M., & Schwartz, B. (2011). Too much of a good thing: The challenge and opportunity of the inverted U. *Perspectives on Psychological Science*, 6, 61–76. [\[CrossRef\]](#)
- Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry*, 26(1), 1–26. [\[CrossRef\]](#)
- Gutentag, T., Halperin, E., Porat, R., Bigman, Y. E., & Tamir, M. (2017). Successful emotion regulation requires both conviction and skill: Beliefs about the controllability of emotions, reappraisal, and regulation success. *Cognition and Emotion*, 31(6), 1225–1233. [\[CrossRef\]](#)
- Hanushek, E. A. (2011). Valuing teachers: How much is a good teacher worth. *Education Next*, 11(3), 40–45.
- Harley, J. M., Lajoie, S. P., Frasson, C., & Hall, N. C. (2017). Developing emotion-aware, advanced learning technologies: A taxonomy of approaches and features. *International Journal of Artificial Intelligence in Education*, 27, 268–297. [\[CrossRef\]](#)
- Hojtink, H., Mulder, J., van Lissa, C., & Gu, X. (2019). A tutorial on testing hypotheses using the Bayes factor. *Psychological Methods*, 24(5), 539–556. [\[CrossRef\]](#) [\[PubMed\]](#)
- Holbert, N. (2020). Constructionism as a pedagogy of disrespect. In N. Holbert, M. Berland, & Y. B. Kafai (Eds.), *Designing constructionist futures: The art, theory, and practice of learning designs* (pp. 141–149). MIT Press.
- Joseph, D. L., Chan, M. Y., Heintzelman, S. J., Tay, L., Diener, E., & Scotney, V. S. (2020). The manipulation of affect: A meta-analysis of affect induction procedures. *Psychological Bulletin*, 146(4), 355–375. [\[CrossRef\]](#)
- Kalokerinos, E. K., Greenaway, K. H., Pedder, D. J., & Margetts, E. A. (2014). Don't grin when you win: The social costs of positive emotion expression in performance situations. *Emotion*, 14(1), 180–186. [\[CrossRef\]](#)
- Kapur, M., & Bielaczyc, K. (2012). Designing for Productive Failure. *Journal of the Learning Sciences*, 21(1), 45–83. [\[CrossRef\]](#)
- Kim, C., & Pekrun, R. (2014). Emotions and Motivation in Learning and Performance. In J. Spector, M. Merrill, J. Elen, & M. Bishop (Eds.), *Handbook of research on educational communications and technology*. Springer. [\[CrossRef\]](#)
- Knol, M. H., Dolan, C. V., Mellenbergh, G. J., & van der Maas, H. L. (2016). Measuring the quality of university lectures: Development and validation of the instructional skills questionnaire (ISQ). *PLoS ONE*, 11(2), e0149163. [\[CrossRef\]](#)
- Koole, S. L. (2009). The psychology of emotion regulation: An integrative review. *Cognition and Emotion*, 23(1), 4–41. [\[CrossRef\]](#)

- 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, 101785. [\[CrossRef\]](#)
- Leach, C. W., & Cidam, A. (2015). When is shame linked to constructive approach orientation? A meta-analysis. *Journal of Personality and Social Psychology*, 109(6), 983–1002. [\[CrossRef\]](#)
- Lench, H. C., Reed, N. T., George, T., Kaiser, K. A., & North, S. G. (2024). Anger has benefits for attaining goals. *Journal of Personality and Social Psychology*, 126(4), 587–602. [\[CrossRef\]](#) [\[PubMed\]](#)
- Loibl, K., Roll, I., & Rummel, N. (2017). Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review*, 29(4), 693–715. [\[CrossRef\]](#)
- Mauss, I. B., Tamir, M., Anderson, C. L., & Savino, N. S. (2011). Can seeking happiness make people unhappy? Paradoxical effects of valuing happiness. *Emotion*, 11(4), 807–815. [\[CrossRef\]](#)
- Meyer, D. K., & Turner, J. C. (2007). Scaffolding emotions in classrooms. In P. A. Schutz, & R. Pekrun (Eds.), *Emotion in education* (pp. 243–258). Elsevier Academic Press. [\[CrossRef\]](#)
- Naismith, L. M., & Lajoie, S. P. (2018). Motivation and emotion predict medical students' attention to computer-based feedback. *Advances in Health Sciences Education*, 23, 465–485. [\[CrossRef\]](#) [\[PubMed\]](#)
- Naragon-Gainey, K., McMahon, T. P., & Chacko, T. P. (2017). The structure of common emotion regulation strategies: A meta-analytic examination. *Psychological Bulletin*, 143(4), 384–427. [\[CrossRef\]](#)
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18, 315–341. [\[CrossRef\]](#)
- Quoidbach, J., Mikolajczak, M., & Gross, J. J. (2015). Positive interventions: An emotion regulation perspective. *Psychological Bulletin*, 141(3), 655–693. [\[CrossRef\]](#)
- Rosiek, J. (2003). Emotional scaffolding: An exploration of the teacher knowledge at the intersection of student emotion and the subject matter. *Journal of Teacher Education*, 54(5), 399–412. [\[CrossRef\]](#)
- Scheier, M. F., & Carver, C. S. (1985). Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Health Psychology*, 4(3), 219–247. [\[CrossRef\]](#) [\[PubMed\]](#)
- Seligman, M. E. P. (1991). *Learned optimism: How to change your mind and your life*. Pocket Books.
- Sinha, T. (2022). Enriching problem-solving followed by instruction with explanatory accounts of emotions. *Journal of the Learning Sciences*, 31(2), 151–198. [\[CrossRef\]](#)
- Sinha, T. (2025). Emotion regulation during failure-driven learning. In *Proceedings of the 19th international conference of the learning sciences—ICLS 2025*. International Society of the Learning Sciences.
- Sinha, T., & Dhandhanian, S. (2022). Democratizing emotion research in learning sciences. In *International conference on artificial intelligence in education* (pp. 156–162). Springer International Publishing. [\[CrossRef\]](#)
- Sinha, T., & Kapur, M. (2021a). Robust effects of the efficacy of explicit failure-driven scaffolding in problem-solving prior to instruction: A replication and extension. *Learning and Instruction*, 75, 101488. [\[CrossRef\]](#)
- Sinha, T., & Kapur, M. (2021b). When problem solving followed by instruction works: Evidence for productive failure. *Review of Educational Research*, 91(5), 761–798. [\[CrossRef\]](#)
- Sinha, T., Kapur, M., West, R., Catasta, M., Hauswirth, M., & Trninic, D. (2021). Differential benefits of explicit failure-driven and success-driven scaffolding in problem-solving prior to instruction. *Journal of Educational Psychology*, 113(3), 530–555. [\[CrossRef\]](#)
- Tamir, M. (2016). Why do people regulate their emotions? A taxonomy of motives in emotion regulation. *Personality and Social Psychology Review*, 20(3), 199–222. [\[CrossRef\]](#)
- Tamir, M., Bigman, Y. E., Rhodes, E., Salerno, J., & Schreier, J. (2015). An expectancy-value model of emotion regulation: Implications for motivation, emotional experience, and decision making. *Emotion*, 15(1), 90–103. [\[CrossRef\]](#)
- Tangney, J. P., & Tracy, J. L. (2012). Self-conscious emotions. In M. R. Leary, & J. P. Tangney (Eds.), *Handbook of self and identity* (2nd ed., pp. 446–478). Guilford Press.
- Tracy, J. L., & Robins, R. W. (2006). Appraisal Antecedents of Shame and Guilt: Support for a Theoretical Model. *Personality and Social Psychology Bulletin*, 32(10), 1339–1351. [\[CrossRef\]](#)
- Travis, J., Kaszycki, A., Geden, M., & Bunde, J. (2020). Some stress is good stress: The challenge-hindrance framework, academic self-efficacy, and academic outcomes. *Journal of Educational Psychology*, 112(8), 1632–1643. [\[CrossRef\]](#)
- Turner, J. E., & Schallert, D. L. (2001). Expectancy–value relationships of shame reactions and shame resiliency. *Journal of Educational Psychology*, 93(2), 320–329. [\[CrossRef\]](#)
- Vallerand, R. J., Chichekian, T., Verner-Filion, J., & Bélanger, J. J. (2023). The two faces of persistence: How harmonious and obsessive passion shape goal pursuit. *Motivation Science*, 9(3), 175–192. [\[CrossRef\]](#)
- Wang, F., Sinha, T., & Kapur, M. (2024). Which bright side to look at? Cultivating goal-oriented shame regulation under challenging problem-solving contexts. In *Proceedings of the 18th international conference of the learning sciences—ICLS 2024* (pp. 226–233). International Society of the Learning Sciences. [\[CrossRef\]](#)

- Weidman, A. C., & Kross, E. (2021). Examining emotional tool use in daily life. *Journal of Personality and Social Psychology*, 120(5), 1344–1366. [[CrossRef](#)] [[PubMed](#)]
- Willroth, E. C., Young, G., Tamir, M., & Mauss, I. B. (2023). Judging emotions as good or bad: Individual differences and associations with psychological health. *Emotion*, 23(7), 1876–1890. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.