Enriching problem-solving followed by instruction with explanatory accounts of emotions

Tanmay Sinha

Professorship for Learning Sciences and Higher Education, Eth Zürich

ABSTRACT

Background: Problem-solving followed by instruction (PS-I) is a powerful design shown to transform students’ conceptual understanding and transfer. Within PS-I, no research has examined how moment-by-moment determinants of affective states impact the problem-solving phase and posttest performance.

Methods: I develop a multimodal learning analytics pipeline to (a) infer affective states in PS-I via observable facial movements, (b) understand how the incidence and temporal dynamics of these states vary based on manipulating the problem-solving context with scaffolding strategies (failure-driven, success-driven, none) in an experimental study (N = 132), and (c) assess the extent to which affective states might explain learning.

Findings: Students exposed to failure-driven scaffolding show exclusive dynamics comprising shame, a self-conscious emotion associated with metacognitive and cognitive benefits. Failure-driven scaffolding also creates opportunities for relatively greater emotional displays of knowledge emotions (e.g., surprise, interest). Hostile emotions differentially impact learning in PS-I, with the incidence of anger and disgust showing positive associations and the incidence of contempt showing a negative association. Finally, pleasurable emotions (e.g., happiness) positively associate with isomorphic posttest performance but negatively associate with non-isomorphic and transfer posttests.

Contribution: Overt changes in facial movements reflective of students experiencing negative emotional states act as catalysts for learning.

CONTACT

Tanmay Sinha tanmay.sinha@gess.ethz.ch Professorship for Learning Sciences and Higher Education, Eth Zürich, RZ J 5, Clausiusstrasse 59, 8092, Switzerland.

© 2021 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Since the seminal work on preparation for future learning (Schwartz & Bransford, 1998), educational research has taken great strides in testing approaches that afford students sensemaking opportunities before formal instruction (see Sinha and Kapur (2021b) and Sinha and Kapur (2021c) for meta-analyses of the field). Preparatory problem-solving is one such sense-making activity that is posited to activate students’ relevant prior knowledge, increase awareness of knowledge gaps, and their desire to know the canonical answer (Loibl et al., 2017). Converging empirical evidence in diverse task domains shows that problem-solving followed by instruction (PS-I) improves students’ conceptual understanding and transfer, in comparison to instruction-first approaches. For example, the meta-analysis by Sinha and Kapur (2021b) suggests a Hedge’s g 0.36 [95% CI 0.20; 0.51] favoring PS-I (N =166 comparisons).

However, despite the growing effectiveness of PS-I, there is no evidence of the extent to which students’ emotions during the process of problem-solving impact learning. With a steadily growing body of literature looking into the explanatory basis for why PS-I designs work (Kapur, 2016; Loibl et al., 2017; Sinha & Kapur, 2021b), there is a need for more fine-grained analysis of the temporal dynamics of cognition and affect. Previous research has implicated that students’ affective disposition might be a critical predictor of their entry into STEM disciplines, even more so than enrollment and grades (Maltese & Tai, 2011). More generally, with mounting evidence on the importance of social, emotional, and cognitive development for a host of positive outcomes such as academic achievement, well-being, career and economic stability (Jones et al., 2019), it is necessary to explicitly account for students’ affect in educational interventions. Given the critical role of emotional processes in learning, as I will further argue in the article, the theoretical architecture of PS-I therefore needs to be expanded to holistically consider both cognitive and affective aspects of learning.

Most learning contexts focusing on fine-grained assessments of natural expressions of discrete affect, however, comprise students who (a) have either already formally learned a concept and are only practicing problems as a follow-up, and/or, (b) are undertaking a guided discovery learning module with success-driven scaffolding nudging them toward the canonical answer. Examples of such contexts include one-on-one sessions with intelligent tutoring systems, interaction with simulations and serious games, and the use of basic computer interfaces for practicing for standardized tests, reading and writing (D’Mello, 2013; Loderer et al., 2018). In interacting with these learning technologies on tasks with low to moderate urgency and low severity of failure consequences, unsurprisingly, meta-analytic evidence has found surprise and negative affective states like contempt, anger, disgust, sadness, and fear to be consistently infrequent (D’Mello, 2013).
PS-I, where the problem-solving phase is a preparation for learning from instruction and not focused on practice with already learned concepts, offers a unique learning context to assess the generalizability of these results. Because the initial problem-solving phase is naturally designed to be ill-structured and afford the generation of multiple suboptimal solutions (Kapur & Bielaczyc, 2012), it is plausible to expect salient differences in the distribution of discrete affect categories in PS-I. The design of explicit failure-driven scaffolding to nudge students to generate suboptimal solutions (Sinha et al., 2021) may further aggravate experiences of negative affect. This, however, may not necessarily “have crippling effects on task persistence, motivation, and learning gains” (D’Mello, 2013, p. 1094), as implicated in previous work. Concerning the evaluation of learning, the inclusion of an explicit formal instruction phase in PS-I affords to differentiate performance during the problem-solving phase and the posttests administered after instruction. This also contrasts with prior work, where learning is often narrowly construed based on the interaction between the student and the learning technology, usually independent of formal classroom instruction.

Further, although the theory underlying PS-I (Loibl et al., 2017) comprises a process-focused explanatory basis (e.g., students activating prior knowledge to typically generate failed solutions, realizing their knowledge gaps in the process and becoming motivated to learn the canonical concept), the learning mechanisms explored thus far rely primarily on (a) retrospective self-report questionnaires, and/or (b) aggregated indicators of problem-solving process, that is, quantity and quality of student-generated solutions. For instance, in a typical PS-I design, task experience surveys that tap onto different cognitive and/or affective mechanisms are administered post-hoc, and students retrospectively report their consciously held beliefs, for example, working through the problem made clear that I lack some knowledge, raised my interest, etc. Such feelings tend to be subjective reactions to the different emotions students experience when they engage in preparatory problem-solving, and are shaped by personal interpretations and reflections.

Although such coarse-grained data about students’ learning offers valuable insights into the explanatory basis of PS-I, it does not sufficiently describe the complexity of learning processes that can be captured by fine-grained behavioral responses (e.g., emotions) and their temporal dependencies. During problem-solving prior to instruction, several automatic behavioral responses might occur not entirely within students’ conscious awareness and are therefore relatively more difficult to control/manipulate. Such responses may offer valuable insights into students’ actual task experiences and decision-making that result in differential learning outcomes. Emotions, which are posited to precede feelings and are characterized by intense mental activity and a high degree of pleasure or displeasure (Cabanac, 2002), are examples of such behavioral responses. Facial
movements are one of many correlates of emotion (Keltner et al., 2019), and they serve as a physical cue to measure and infer affective states as students engage in the problem-solving activity or attend to instruction. The time scale of facial movements affords relating them to learning events in which they occur.

Taken together, the aforementioned process-focused nature of PS-I’s theoretical architecture and the rather crude methods used to evaluate empirical evidence for the theory create an apparent divide. This gap in research seems to be (a) theoretically-rooted, evidenced via the majority of PS-I research foregrounding cognitive learning mechanisms, as well as (b) methodologically-rooted, evidenced via lack of research on implementing multimodal learning analytics to study fine-grained problem-solving and affective processes. To fill these research gaps, I propose an approach grounded in advances in affective computing, with the objectives to (a) discover moment-by-moment determinants of students’ affective states during preparatory problem-solving, (b) understand how the incidence and temporal dynamics of these states vary based on manipulating the problem-solving context with scaffolding strategies (failure-driven, success-driven, none), and (c) assess the extent to which affective states might explain learning. See Figure 1 for a summary. Specifically,

![Figure 1](image-url)
(1) I develop a fully automated pipeline that can
(a) process recorded videos of students using a standard webcam in a PS-I design.
(b) leverage algorithms from state-of-the-art open-source tools to detect facial movements and head motion with high reliability.
(c) apply a set of coding rules based on multiple psychological lenses for emotion inference from facial movements on a frame-by-frame basis (1/30th of a second).
(d) compute the incidence of inferred emotional states (percentage of video frames coded into a particular emotion category), and finally,
(e) using big data statistical procedures, discover emotion dynamics, that is, frequently co-occurring and temporally contingent sequences of emotions over varying periods.

(2) I offer interpretations for the emotional states with high incidence and/or those that are a part of frequently occurring behavioral sequences by
(a) contextualizing results in the light of self-reported affective task experiences, and computing correlations with problem-solving phase and posttest performance.
(b) drawing on the rich body of literature that links cognition and emotion, and describes the psychological functions that emotions serve.

Theoretical background

A failure-first instructional approach

Because my apparatus for studying emotions is situated in a scaffolded PS-I design foregrounding learning from failure, I first highlight the underlying instructional design rationale. This line of work builds on the long-standing debate about whether scaffolded problem-solving experiences should lean toward being more success-driven or failure-driven (Kapur, 2016). Existing empirical studies that actively elicit or induce failure mainly rely on adding challenges, for example, via forced guessing—to the best of my knowledge, research on explicitly scaffolding problem-solving toward failure by nudging students to reason with suboptimal representations, or deliberate guided failure is non-existent in cognitive and educational psychology (Sinha et al., 2021). Taking inspiration from structuring versus problematizing-focused accounts of scaffolding that either decrease or increase the degrees of freedom in a task (Reiser, 2004), I have therefore previously designed a classroom study (Sinha
et al., 2021) and a follow-up lab study (Sinha & Kapur, 2021a) to evaluate the pedagogical value of failure-driven and success-driven scaffolding in preparatory problem-solving.

Empirical results across these studies suggest that nudging students to generate suboptimal representations (a) affords exploration of the problem-space, (b) provides variable practice, and (c) traps potential misconceptions, despite challenging students’ understanding and making them uncomfortable in the short-term. These posited benefits of explicit failure-driven exploration have been manifested in relatively higher scores and better reasoning quality at the posttest for non-isomorphic and transfer questions, relative to success-driven and/or an unscaffolded condition. By applying and evaluating the effect of a suggested suboptimal representation that should naturally result in a misprediction, students have opportunities to see the limitations of that representation, and become more aware of the boundary conditions of that prediction. In that sense, this line of work can also be conceived to be in the spirit of trapping misconceptions, or using such mispredictions to destabilize suboptimal representations during a follow-up lecture.

This means that the instructional design rationale shifts the focus from providing known evidence and emphasizing why right is right to confronting students with conflicting evidence and emphasizing negative knowledge, or why wrong is wrong. Taken together, the efficacy of generation, coupled with nudging students to generate specific suboptimal representations can be posited to contribute to improved understanding of the targeted learning concept. An intuitive and rather well-practiced alternative is to simply tell students why wrong is wrong. Decades of research on conceptual change (Vosniadou, 2009), however, suggest that students’ suboptimal beliefs about the efficacy of problem-solving representations are resistant to direct instruction.

The extent to which implementation of a failure-first instructional approach is an ethical scaffolding strategy depends largely on how learning activities foregrounding failure are framed, and whether there exists a positive and trustworthy climate in the classroom to facilitate appreciation for failure-driven learning. Along with providing empathy concerning possible frustration of students, a positive classroom climate is also likely to increase their awareness of the pedagogical benefits of engaging in errorful generation. In the aforementioned studies (Sinha & Kapur, 2021a; Sinha et al., 2021) for instance, the (a) problem-solving activity was explicitly situated in authentic interdisciplinary domain practices upfront via the information sheet participants studied prior to providing consent, (b) appropriate socio-mathematical norms were enforced, for example, by setting the expectation that multiple solutions were expected and failure was acceptable and valued, and (c) motivational prompts were used to keep students engaged in generation via multiple task announcements.
Against this backdrop of a failure-first instructional approach as the study context for the article’s focus on the interplay of cognition and affect in problem-solving, I now go on to describe relevant research that motivates why facial expression analyses might offer a valid/meaningful account of emotions. Following that, I describe dominant theoretical and methodological approaches to studying emotions in learning, with the goals of emphasizing (a) the contrasting perspectives in the literature on the learning effects of positive versus negative emotions, (b) limitations in prior methodologies that motivate the reported affective computing approach, and tying together (c) the learning mechanisms from previously reported studies.

**Emotions, cognition and learning**

Research in affective neuroscience posits that the changes in body and brain states in response to stimuli, also referred to as somatic markers (Damasio, 1994), influence subsequent decision-making. The facial feedback hypothesis (Coles et al., 2019), which focuses on one particular sub-category of somatic markers suggests that an individual’s experience or sensation of emotion is influenced by sensorimotor feedback from selective activation and inhibition of facial muscles, prototypically associated with the expression of emotion.

The potential for experienced emotions to be masked and not socially expressed (Cahour, 2013) might explain why most studies in the psychology of emotions are developed through experimental settings. Here, usually contrasting emotional states are induced before observing differences in subsequent behavior. For example, to enhance learning outcomes, emotional design manipulates affective qualities of the environment such as colors, shapes, expressions of characters and sounds. However, recent meta-analytic evidence from this experimental paradigm suggests that compared to elementary, middle and high school students, emotional design manipulations administered for postsecondary students (the population of the current study) have much smaller effects on learning and motivational outcomes (Wong & Adesope, 2020). Therefore, in contrast, here I capture raw, unfiltered, and naturally-occurring emotional responses in an authentic learning task, without superficially changing features inherent to the problem-solving environment but detached from the learning context. I expect students’ emotions to be more strongly influenced by how they appraise sensemaking activities conceptually relevant to the problem-solving task, that is, by engaging in cognitive processing germane (Sweller, 2011) to learning.

On the other hand, meta-analytic evidence from an alternative predominantly observational research paradigm suggests that the majority of the work on studying moment-to-moment emotions during learning for postsecondary students falls into domains such as computer literacy, creative/argumentative writing, and research methods (D’Mello, 2013). To the best of
my knowledge, there is no research on studying moment-by-moment emotions within the domain of data science education (the domain of the current study), in part also owing to the fairly recent development of data science as a field of study (Vahey et al., 2017; Wilkerson & Polman, 2020). More broadly, despite an increasing number of multimodal research studies being conducted in university settings with undergraduate students as the target group (from 2000–2017), the dominant data collection modality comprises interviews, surveys, and audio/video observations focused primarily on measuring the cognitive aspect of learning. For instance, only 11% of all multimodal research studies included in a recent review of the field studied the emotional aspects of learning (Noroozi et al., 2020).

In summary, there exists a research gap in investigating the interplay between cognition and fine-grained affect for instructional approaches following a problem-solving first pedagogy, especially via the advances in the reported affective computing methods. Given the growing effectiveness of PS-I approaches (Sinha & Kapur, 2021b, 2021c), it is important to understand and triangulate evidence for underlying affective and cognitive mechanisms that drive learning benefits. This article’s focus on a fine-grained investigation into emotional states is one way of approaching this important theoretical question.

**Theoretical lenses**

I take a socio-constructivist theoretical grounding on learning, where novices come to make sense of a situation by first acting on it and then receiving expert guidance (Vygotsky, 1987). The participation metaphor (Sfard, 1998) suggests learning to be contingent on having a shared understanding of the world and implicates that as novices, students cannot be expected to see what experts see. This is because seeing is a function of what one knows, and novices do not have the necessary prior knowledge and experience of the experts (Chi et al., 2014). Hence, novices are not always prepared to take on the opportunity to hear and practice canonical ideas endorsed by experts, even when delivered in a clear and well-structured manner (Schoenfeld, 1988). Engaging novices in preparatory problem-solving activities that help activating and differentiating prior knowledge in relation to the targeted concept is one route that can help.

Within this preparatory problem-solving framework and also more generally, I consider emotion and motivation as an intrinsic part of the design and analysis of learning. This means that I view the link between cognition and affect as reciprocal, or as dynamically interacting systems that strengthen and weaken each other in a multitude of ways. This has direct implications for the measurements I have deployed within the PS-I learning design in past work (Sinha & Kapur, 2021a; Sinha et al., 2021). These measurements focus both on (a) cognitive engagement, for example, how invested students are in
a task, their willingness to exert effort, and (b) emotional engagement, for example, students’ attitudes about the learning task, including emotional displays and subjectively expressed feelings that they can engage.

Given that I view affective states as causes and catalysts of learning, here I begin by reviewing two influential theoretical lenses that explain the antecedents and effects of emotions experienced by students—(a) the cognitive-affective theory of learning with media (Moreno & Mayer, 2007), which emphasizes the mediating effect of affect, emotions, motivations, etc on the cognitive engagement of students, that is, their processing of visual and verbal information, and (b) the control-value theory of achievement emotion (Loderer et al., 2018; Pekrun & Perry, 2014), which describes the differential influence of positive emotions (better for learning) and negative emotions (relatively worse for learning) on achievement outcomes. However, a broader review of the psychological literature implicates an opposing view, especially in perspectives from motivation (Kashdan & Biswas-Diener, 2014) and second-wave positive psychology (Ivtzan et al., 2015). These strands view negative emotions as indicators of challenging stimuli that motivate actionable work forward, and should therefore be embraced.

Taken together, after the consolidation of evidence from the aforementioned theoretical lenses, I adopt a balanced approach to negative emotions when stumbling—allowing oneself to feel bad, but not letting such emotions take over. My take is that negative emotions hold the potential to engage students in germane processing of task information (Sweller, 2011), and act as a critical juncture signaling disequilibrium in the learning process. Therefore, experiencing moderate levels of negative emotions triggered by the design affordances of PS-I is all right, as long as it is appraised appropriately, for example, as a challenge rather than a threat. Not all negative emotional responses are bad, especially if they facilitate preparatory task goals.

Further, it is also important to note that although prior theoretical lenses focus on a broad set of academic emotions that are relevant in educational contexts such as technology-based learning environments, the dominant research methodology to analyze affect comprises psychometrically grounded surveys tapping into (a) positive and negative affective states directly, for example, via PANAS scale (Watson et al., 1988) or similar, and/or, (b) factors such as goals, situational interest and self-concept that might serve as a precursor to the experience of an affective state. As a result of this exclusive focus on a coarse-grained sampling of affect, there is a risk of missing the contingencies between different affect categories across a finer-grained temporal resolution. Here, I therefore adopt a convergent parallel mixed-methods approach to combine self-reports of students’ affective task experiences with the study of fine-grained emotion dynamics and relate these process measures to learning outcomes.
Empirical work

To adjudicate between differential views regarding the impact of positive and negative emotions on learning, empirical work has begun to identify conditions under which such differential effects might hold. For example, in learning with technology, the incidence of frustration (dissatisfaction or annoyance) and boredom (being weary or restless through lack of interest) has been found to be related to lower learning outcomes compared to students who did not experience these emotions (D’Mello & Graesser, 2012). However, students who experienced confusion (a noticeable lack of understanding) were found to score high on transfer assessments (D’Mello et al., 2014). Here, emotional states were captured via a retrospective affect judgment protocol following the posttest—students watched their own videos and paused every few seconds to choose from explicitly defined emotional states. In contrast, I refrain from relying on students’ ability to consciously access their emotional states—a procedure that is likely to be heavily influenced by inaccurate recall, the desire to be honest and/or other social biases, for example, the unwillingness to report negative emotions like disgust. Instead, I use (a) culturally-generalizable emotion inference protocols (Cordaro et al., 2018; Du et al., 2014; Keltner et al., 2019), and (b) a categorization scheme going beyond the simplistic dichotomy of positive and negative emotions, which organizes the inferred emotions into categories like self-conscious, knowledge, hostile and pleasurable (Silvia, 2009) that more accurately depict the psychological functions served by emotions.

For instance, self-conscious emotions like shame require evaluation as to whether one has lived up to a given standard or goal and can increase achievement strivings by motivating an approach-orientation toward failure (Leach & Cidam, 2015), especially when task situations are perceived as controllable. PS-I situations with high failure-likelihood exemplify such task situations because of the presence of constant situational feedback from the problem-solving environment including negative knowledge from scaffolds, especially if designed to explicitly nudge students toward failure. On the other hand, when chances of success are high, emotional experiences of shame are plausibly expected to be low because students’ ability and/or effort are less likely to stifle task progress.

In contrast to self-conscious emotions that are important but understudied in PS-I, knowledge emotions like interest, surprise and confusion are more closely tied to previously studied learning mechanisms posited to underlie the effectiveness of PS-I. For instance, Loibl and Rummel (2014) advance the argument that students’ awareness of knowledge gaps that surface when generating multiple solutions during the problem-solving phase of PS-I might raise their interest to know the canonical answer. PS-I implementations that trigger a higher intensity of knowledge gap awareness may therefore be expected to have a higher intensity of emotional displays of
interest. In contrast to unscaffolded PS-I, in which students work with their own ideas and what they know, scaffolding PS-I toward failure or success by providing specific (sub)optimal representations to generate is likely to provide additional sensemaking opportunities, and thus relatively greater opportunities for raising knowledge gap awareness about solution strategies that may not work. In fact, self-reported data for the current sample reported in Sinha and Kapur (2021a), where the difference in scores for students’ knowledge gap awareness between the scaffolded and unscaffolded PS-I conditions was high (Cohen’s $d > 0.5$), empirically supports this conjecture. Further, irrespective of whether scaffolds decrease or increase the degrees of freedom in the task, they present new information that adds to the already unpredictable nature of the problem-solving phase because of the lack of verifiable outcomes. Applying and evaluating the effect of such novel problem-solving scaffolds is therefore likely to intensify emotional displays of surprise generated by the expectancy-violating events that follow. Finally, emotional displays of confusion, a dominant behavioral signature of cognitive disequilibrium in problem-based learning (D’Mello et al., 2014) can be expected to be widely prevalent in PS-I, however, with a relatively less pronounced incidence in the presence of success-driven scaffolds that provide students more meaningful ways to progress toward the solution.

The third category of emotions for which there is plausibly a strong theoretical rationale concerning their role in PS-I’s effectiveness but only partial and/or indirect empirical support is hostile emotions, in particular anger, disgust, and contempt. I argue that these emotions lie at the heart of the underlying pedagogical rationale of PS-I, that is, placing students into a state of dissonance by focusing their attention on recognizing deep features of the task situation and implicitly illustrating ways in which their understanding may be flawed (Kapur & Bielaczyc, 2012; Loibl et al., 2017; Sinha & Kapur, 2021b). Additionally, scaffolds nudging problem-solving toward failure provide conflicting evidence with respect to previously generated inferences about the task solution (Sinha et al., 2021). Taken together, this buildup of an uncomfortable tension across the problem-solving phase is likely to result in emotional displays comprising hostile emotions. Although anger can be perceived as unprovoked or inexplicable, the basic psychological dynamic at work is that it is a response to what is perceived as unfair (Roseman, 1991). In contrast to problem-solving anxiety that is associated with future uncertainty and the inability to control what will happen, experiences of anger within PS-I situations are likely to encourage active behaviors to address problematic aspects and mollify the task situation (Harmon-Jones et al., 2014). Such an approach also aligns with the theory of cognitive dissonance (Festinger, 1962), which would predict students to be subsequently motivated to engage in problem-solving strategies that probe into the problematized subject matter, withdraw task resources from those
actions that do not show promise and redeploy to those that do. Similar to anger, although emotional experiences of disgust may stem from psychologically different roots (Miller, 1998), they are confrontational and trigger change and should therefore positively affect learning. However, contempt reflects an attitude of indifference toward the problem-solving situation (Fischer & Giner-Sorolla, 2016) and its incidence is likely to be negatively associated with learning outcomes. Further, in PS-I implementations where the degrees of freedom in the task are reduced and problem-solving is made relatively more tractable via success-driven scaffolds, triggers for emotional displays of hostile emotions are expected to be low and therefore such emotions may have a negligible association with learning outcomes.

Finally, emotional displays falling into the category of pleasurable emotions (e.g., happiness) can serve as a double-edged sword. The positive association between subjective experiences of positive affect and learning outcomes in previous empirical PS-I work (Lamnina & Chase, 2019) suggests that, on the one hand, happiness can reflect a satisfactory response tendency toward problem-solving progress. Such chances are higher in isomorphic problem-solving at the posttest, and especially for students exposed to success-driven scaffolds. However, quick or easy success in the problem-solving phase has the potential to set students up for failure by providing insufficiently disruptive problem-solving experiences that do not challenge existing thought processes enough. Therefore, happiness may also alternatively index complacency and may co-exist (happily!) with students not realizing the faults they have to work on, in turn leading to reduced learning from instruction. Irrespective of whether PS-I is scaffolded or not, this can have detrimental learning effects when students solve posttest questions that are non-isomorphic to the problem-solving phase.

More generally, in addition to the aforementioned emotion categories, previous research has implicated emotions like stress and fear to more likely occur in problem-solving situations with high failure likelihood that can tax or exceed an individual’s working memory capacity (LePine et al., 2004; Puttwain et al., 2016). In mathematics problem-solving, there is also some empirical evidence of students with a mastery-goal orientation experiencing positive emotions like interest, enjoyment, and pride after explicit feedback on experiences of failure (Tulis & Ainley, 2011). However, this strand of research has predominantly focused on coarse-grained self-reported emotions, which, as I have described earlier, has its tradeoffs. Further, this empirical work was not carried out in a preparatory problem-solving context. Within the context of failure-driven or success-driven preparatory problem-solving, more research is therefore needed on studying how fine-grained dynamics of negative along with positive emotions play out in real-time and impact learning from instruction. Additionally, although looking at the frequency of occurrence of emotional states in and of itself provides
valuable information for understanding students’ problem-solving experiences, the addition of *temporal information* allows deciphering the meaning of frequently subtle facial expressions that may not be identifiable in static presentations.

**Conceptualization and inference of emotions**

Given that emotions and cognition are deeply interconnected (Lerner et al., 2015; MacCann et al., 2020), natural follow-up questions are how to best conceptualize emotions and perform emotion inference. Although several theoretical perspectives have been put forth to *conceptualize emotions*, the distinction between *discrete* (Ekman & Friesen, 1976) and *dimensional* (Russell, 1980) models is worth pointing out. Discrete emotion theorists acknowledge a set of emotion categories that are fundamentally different (e.g., happiness, sadness, surprise), each with different physiological processes, accompanying cognitions, and observable behavioral signatures (e.g., facial movements). Dimensional models emphasize grouping and arranging these discrete emotion categories into more primitive dimensions of valence (positivity versus negativity) and arousal (activating versus calming).

With respect to *emotion inference*, there is again a distinction between the *common view* (Ekman et al., 1987) versus *appraisal-based* (Ellsworth & Scherer, 2003) scientific frameworks. The common view posits that each emotion is reliably signaled and revealed with a prototypical configuration of facial movements, which vary to some degree depending on processes that are independent of an emotion itself (e.g., emotion regulation strategies and culture-specific dialects). The appraisal perspective posits the existence of a larger set of non-prototypical emotions, whose expressions vary based on the evaluation of the context.

**The present analysis**

Both discrete and dimensional perspectives for conceptualizing emotions have much to offer in attempting to understand the psychology of emotions, despite not perfectly modeling the underlying emotion-generation processes. The fact that emotions can be described in terms of dimensions does not negate the importance of a discrete emotions perspective. For example, contempt differs qualitatively from anger, despite both being negatively valenced. Additionally, the relatively finer distinction between emotion categories in discrete models, makes them more suited for the design of action-able affective response strategies via a teacher or an automated coach in a learning context. In the present study, my analysis therefore aligns with the *discrete* conceptualization of emotion.
In a similar vein, based on a rich body of prior empirical work describing facial movements in learning (see Cahour (2013) for a review)—work, which has been heavily influenced by the common view theoretical framework—my analysis too aligns more strongly with this emotion inference framework. Based on the recommendations in Barrett et al. (2019), I, however, attempt to overcome some underlying criticisms, by:

1. not restricting the set of discrete emotion categories to six universal ones (i.e., anger, disgust, fear, happiness, sadness, and surprise). I include additional non-prototypical emotions that may be relevant in a learning context (e.g., shame, interest, confusion).

2. accounting for an individual’s expression of emotional states that cannot be assigned to only one distinct discrete category (e.g., happy, angry, surprised). I also examine facial movements reflecting such compound emotions (e.g., happily surprised, angrily surprised).

3. acknowledging that cultural aspects can be essential for the classification of expressed emotions. I, therefore, use three culturally generalizable coding schemes borrowed from state-of-the-art research in affective sciences (Cordaro et al., 2018; Du et al., 2014; Keltner et al., 2019) to perform emotion inference from facial movements, and integrate the findings for my data. Each coding scheme draws on a different theoretical orientation and is based on large-scale human-coded datasets possessing diverse demographics.

4. manipulating the context in which facial movements occur. I randomly assign students to three PS-I designs, with each group receiving either no explicit scaffolding (henceforth referred to as the Productive Failure condition), success-driven scaffolding, or failure-driven scaffolding in the problem-solving phase (Sinha & Kapur, 2021a). This allows evaluating the extent to which emotion inference is influenced by the learning design.

5. not adopting an early fusion approach to down-sample the facial movement data at an adhoc time granularity, with the risk of operating on an insufficient temporal resolution. To maximize the amount of reliable information available for emotion inference, I preserve the facial movement time-series data at its original temporal resolution (i.e., 1/30th of a second). My data mining approach for capturing emotion dynamics therefore operates on 14.8 million video frames.
Research questions and hypotheses

In light of the above-discussed work, the present study addresses the following research questions and their associated hypotheses. Note that for RQ1 and RQ2, hypotheses for particular emotion categories are formulated as explanatory if they are either (a) linked to previously studied PS-I learning mechanisms (e.g., interest, affect, cognitive dissonance), and/or if there (b) exists substantial former cognitive and educational psychology research backing their psychological functions in complex problem-solving (e.g., surprise, shame, confusion, disgust, contempt, pride). On the other hand, for emotions that are (a) currently under-researched in complex problem-solving (e.g., sadness, embarrassment, pain), and (b) for compound emotions that I coded drawing on recent affective computing advances (e.g., happily disgusted, angrily surprised), I conduct exploratory analyses of their role in explaining learning differences and/or their relationship with learning outcomes. For RQ3, except for one explanatory hypothesis as stated below, my analysis of the differences in emotion dynamics across the experimental conditions is predominantly exploratory.

(1) RQ1: How does scaffolding type during preparatory problem-solving (success-driven, failure-driven, none) differentially affect the incidence of inferred emotion categories?
   (a) Hypothesis 1a: Students in the success-driven condition would have the least frequency of overt displays comprising knowledge emotions (e.g., surprise, confusion, interest).
   (b) Hypothesis 1b: Students in the failure-driven and Productive Failure conditions would more often display overt facial movements reflective of hostile emotions (e.g., anger, disgust, contempt), compared to students in the success-driven condition.
   (c) Hypothesis 1c: Students in the failure-driven and Productive Failure condition would more often display overt facial movements reflective of self-conscious emotions (e.g., shame), compared to students in the success-driven condition.
   (d) Hypothesis 1d: Students in the success-driven condition would have the highest frequency of pleasurable emotions (e.g., happiness).

(2) RQ2: How does the incidence of inferred emotion categories correlate with students’ problem-solving phase and posttest performance?
   (a) Hypothesis 2a: Self-conscious emotions (e.g., shame) would be positively associated with learning outcomes in the failure-driven condition and Productive Failure conditions, while they would be unrelated to learning outcomes in the success-driven condition.
(b) Hypothesis 2b: Sub-categories of hostile emotions would be differentially associated with learning outcomes in the failure-driven and Productive Failure conditions, with anger and disgust exhibiting positive correlations, and contempt exhibiting negative correlations. The incidence of hostile emotions would be unrelated to learning outcomes in the success-driven condition.

(c) Hypothesis 2c: Across all the experimental conditions, knowledge emotions (e.g., surprise, confusion, interest) would positively correlate with learning outcomes.

(d) Hypothesis 2d: Pleasurable emotions (e.g., happiness) would be negatively correlated with learning outcomes in the failure-driven and Productive Failure conditions, while they would be positively correlated with learning outcomes in the success-driven condition.

(3) RQ3: What are the trends in frequently occurring emotion dynamics for the problem-solving phase, and how do these vary for students receiving success-driven, failure-driven, or no scaffolding during the preparatory problem-solving activity?

(a) Hypothesis 3a: Students in the failure-driven and Productive Failure conditions would show a wider variation in emotion dynamics, compared to the success-driven condition.

Method

Experimental design

A between-subjects PS-I design (see Figure 2) in the domain of data science education was administered at a public university in Western Europe (N = 132 university participants, novices to the targeted learning concept—59% male, n = 78; 41% female n = 54). Participants were ethnically diverse—58.33% European from ten different ethnicities, 30.3% German being the
majority; 41.67% non-European, 37.12% Asian being the majority. To participate in the study, participants had to know high school math (basic algebra, calculus, statistics and probability) and be familiar with programming in Python (at least 1 semester of Python programming experience). During the problem-solving phase, participants had to create multiple measures to rank-order datasets of four companies from the most successful to the least successful. Information about the number of car units sold and employee satisfaction was given to them. Consistent with the design principles of Productive Failure (Kapur & Bielaczyc, 2012), I designed the datasets such that they had the same non-parametric statistics (median, interquartile range, Spearman’s correlation), but different parametric statistics (mean, standard deviation, Pearson’s correlation) and very different visualizations.

Students in every experimental condition made an identical first attempt at this problem-solving task in the absence of any external scaffolds (20 minutes). For problem-solving in the second attempt, three experimental manipulations were instantiated (20 minutes) and participants were randomly assigned to these conditions. I call these conditions failure-driven \((N = 45,\) explicit problematizing scaffolds pushing students toward failure, or nudging them toward suboptimal solutions), Productive Failure \((N = 43,\) no explicit scaffolds), and success-driven \((N = 44,\) explicit structuring scaffolds pushing students toward success, or nudging them toward optimal solutions). The two scaffolding conditions comprised multi-step scaffolds that were progressively presented. See supplementary materials for more details about the presentation rationale.

Because the preparatory problem-solving task involved reasoning with a bivariate dataset, the failure-driven scaffold hierarchy started with the presentation of (a) moderately high suboptimal representation—one-dimensional histogram, subsequently (b) an extremely high suboptimal representation—bar chart, and finally ended with (c) the least suboptimal representation among the three—two-dimensional histogram. On the other hand, the success-driven scaffold hierarchy included (a) prompt—a Wikipedia page suggesting students to read more about exploratory data analysis, (b) hint—description of data science phenomena under consideration, and finally, (c) bottom-out hint or the last hint in the sequence precisely conveying the answer—syntax for scatterplot generation, an optimal graphical representation for reasoning with bivariate datasets. An implementation fidelity assessment suggested that an adequate number of scaffolds were used \((M = 1.83, SE = 0.14\) for the success-driven condition; \(M = 1.47, SE = 0.13\) for the failure-driven condition). Further, a better response strategy to the scaffolds, reflective of students making changes to the line of reasoning and accommodating information from the scaffolds, was significantly associated with improvements in the generated solution. Finally, for the Productive
Failure condition, students were asked to simply keep generating more solutions for the task. More details about the scaffolding fidelity and study design can be found in Sinha and Kapur (2021a).

The follow-up instruction phase (20 minutes) was presented in the form of a video lecture and kept constant across conditions. The lecture introduced the targeted learning concept, compared and contrasted student solutions during the initial problem-solving task with the canonical solution, and discussed one canonical solution to the task. Finally, students solved posttest questions assessing conceptual understanding and transfer. Before the problem-solving phase, students’ incoming profile was assessed using a combination of questionnaires (e.g., goal orientation, attitude toward mistakes) and a math ability pretest. After the problem-solving phase, I collected students’ task experiences using questionnaires that tap onto different mechanisms (e.g., state curiosity, affect) posited to attribute to the preparatory benefits of PS-I (Sinha et al., 2021). After the posttest, students’ perceived lecture quality was also assessed using questionnaires.

In addition to these retrospective measures, the problem-solving and instruction phases were video recorded with a frontal view of the face and upper body, which is the primary locus of interest here. For each student, I used the Intel RealSense F200 3D webcam integrated into their working stations (Lenovo ideacentre AIO 700–22ISH machines) for performing the video recording at 30 frames per second. The study was approved by the University ethics commission. All subjects gave informed consent to participate in the experiments and agreed to the collection of their video recordings being used to infer cognitive and affective states post-hoc.

**Deriving process measures**

I flesh out the four-phase methodology for deriving process measures of emotion here. See Figure 1 for a summary.

**Coding facial movements (phase one)**

In the first phase, I used the Facial Action Coding System (FACS), which is the de-facto standard for measuring observable changes in facial movements (Ekman & Friesen, 1976), as an objective and reliable basis for my video-based analysis. FACS coding is based on facial action units (AUs) that represent distinct movements displayed on the face, and emerge by activating one or a combination of facial muscles.

Because human observation and coding of facial movements using FACS are time- and labor-intensive, I leveraged advances in automated computer-based analysis (Baltrusaitis et al., 2018) to annotate the presence or absence of activated AUs on a frame-by-frame basis (1/30th of a second). Simply put, facial action unit recognition works by detecting faces, and then significant
regions of interest in the face (e.g., eye corners, nose tip, mouth center). Supervised machine learning algorithms trained using human-coded FACS data serve as the basis for detection. Here, I used recorded videos of the problem-solving and instruction phases as inputs. To maintain high data quality for all downstream analyses, only those video frames were considered that had a face successfully tracked and facial landmarks detected with confidence greater than 80%. This resulted in 5.32 and 5.23 million frames for the first and second problem-solving attempts, and 4.25 million frames for the instruction phase. Camera-placement was pilot tested to avoid any potential bias in facial action unit detection.

**Mapping facial movements to emotions (phase two)**

In the *second phase*, I applied a set of rules based on multiple psychological lenses (Cordaro et al., 2018; Du et al., 2014; Keltner et al., 2019) for emotion inference from the annotated facial movements. I used meta-analytic evidence on the incidence of discrete emotional states during learning with technology (D’Mello, 2013) as a starting point to select the categories of emotion. This led to the inclusion of anger, confusion, contempt, disgust, fear, happiness, interest, and surprise in the first pass. I, however, also expanded on these categories using state-of-the-art evidence from the affective sciences (Cordaro et al., 2018; Du et al., 2014; Keltner et al., 2019). Three coding schemes, rooted in a similar conceptualization of emotion (*discrete*) and emotion inference framework (the *common view*), were applied to highlight a wide palette of emotions students may experience when working through scaffolded problem-solving. The difference in how conservative each scheme was, afforded natural variation in the coding frequency (*incidence*) and thus in the discovered sequential patterns of emotions (*dynamics*). See Table 1 for an overview. I elaborate on these coding schemes below.

(1) The cross-cultural study by Cordaro et al. (2018) was particularly interesting for my analysis because it considered 23 *international core patterns* of how an emotion is expressed using facial movements. Such patterns were found to appear at *above-chance rates* across five distinct cultures varying in their societal characteristics (China, India, Japan, Korea, United States). Further, *no gender differences* were observed in the frequency of occurrence of these patterns across these cultures. Given the demographic diversity of participants in my study, a culturally generalizable coding scheme for emotion inference was plausible. In the final analysis, I included 13/23 emotions. Contentment, relief, triumph, boredom, and sadness were excluded because of not having enough AUs annotated automatically from the first phase. Interest was excluded because it was based on head nod
Table 1. Mapping facial movements to emotions using multiple psychological lenses.

| Emotion       | Action Unit | Physical Description                                                                 |
|---------------|-------------|---------------------------------------------------------------------------------------|
| Amusement     | 6, 7, 12, 25, 26, 53 | Duchenne smile, lower lip dropped, lips parted, jaw dropped, head up                  |
| Happiness     | 6, 7, 12, 25, 26 | Duchenne smile, lower lip dropped, lips parted, jaw dropped                            |
| Awe           | 1, 2, 5, 12, 25, 26, 53 | Eyebrows raised, upper eyelids raised, smile, lips stretched, jaw dropped, head up     |
| Pride         | 7, 12, 53 | Eyelids tight, smile, head back                                                        |
| Surprise      | 1, 2, 5, 25, 26 | Eyebrows raised, upper eyelid raised, lips parted, jaw dropped                         |
| Anger         | 4, 7 | Brows furrowed, upper eyelid raised, eyelids tight, chin raised, lips tightened       |
| Confused      | 4, 7, 55/56 | Brows furrowed, eyelids narrowed, head tilted left/right                                |
| Contempt      | 4, 14, 25 | Brows furrowed, lip corner(s) tight, lips parted                                       |
| Disgust       | 4, 6, 7, 9, 10, 25, 26 | Brows furrowed, outer eyes contracted, eyelids narrowed, nose wrinkled, upper lip raised, chin raised, lips parted, jaw dropped |
| Embarrassment | 6, 7, 12, 25, 54 | Duchenne smile, lips corner depressed, lips parted, head turned right, head down       |
| Fear          | 1, 2, 5, 7, 25, 26, 53 | Brows furrowed, jaws dropped, bodily movement backwards                             |
| Pain          | 4, 6, 7, 20, 25, 53, 54 | -                                                                                   |
| Shame         | 4, 17, 54 | Brows furrowed, lower lip raised, head down                                            |
| Interest      | 1, 2, 12 | Eyebrows raised, lips corner pulled                                                    |
| Sadness       | 1, 4, 6, 15, 17 | Inner eyebrow raised, brows furrowed, cheek raised, lips corner depressed, chin raised |
| Happily       | 1, 2, 5, 12, 25, 26 | Eyebrows raised, upper eyelids raised, smile, lips stretched, jaw dropped              |
| Surprised     | 6, 10, 12, 25 | Cheeks raised, upper lip raised, smile, lips stretched                                 |
| Disgust       | 1, 4, 20, 25 | Inner eyebrow raised, brows furrowed, lips stretched and parted                       |
| Sadly         | 4, 15 | Brows furrowed, lips corner depressed                                                  |
| Sadly         | 1, 4, 25, 26 | Brows furrowed, cheeks raised, upper lip raised                                       |
| Brows furrowed, lips stretched and parted | Brows furrowed, lips stretched and parted |
| Action Unit | Physical Description |
|-------------|----------------------|
| Fearfully   |                      |
| Surprised   |                      |
| Disgusted   |                      |
| Angrily     |                      |
| Surprised   |                      |
| Disgusted   |                      |
| Disgustedly |                      |
| Surprised   |                      |

**Table 1.** (Continued).

| Emotion         | Action Unit | Physical Description |
|-----------------|-------------|----------------------|
| Fearfully       |             |                      |
| Surprised       |             |                      |
| Disgusted       |             |                      |
| Angrily         |             |                      |
| Surprised       |             |                      |
| Disgusted       |             |                      |
| Disgustedly     |             |                      |
| Surprised       |             |                      |
| Appalled/       |             |                      |
| Hatred          |             |                      |

*Note: Table continues with similar entries for other emotions.*
Detection (which involves examining head motion variability across multiple frames) and thus could not be computed frame by frame. Coyness, desire (food), desire (sex), and sympathy were excluded because of non-relevance to the learning scenario.

(2) The review article by Keltner et al. (2019) drew on multiple empirical studies within the common view framework of emotion inference, and synthesized descriptions of 18 emotions with their prototypical facial movements. The coding scheme drawing on this line of work derived prototypical facial-bodily expressions of emotions using data gathered from ten different cultures: China, Japan, Korea, New Zealand, Germany, Poland, Pakistan, India, Turkey, and the United States. In the final analysis, I included 11/18 emotions. Contentment, boredom, pride, shame were excluded because of not having enough AUs annotated automatically from the first phase. Coyness, desire, and sympathy were excluded because of non-relevance to the learning scenario.

(3) The study of compound emotions by Du et al. (2014) was included because of its explicit emphasis on co-occurring emotions that can be *distinctively expressed* because of overlap in AU patterns as well as *unambiguously discriminated* by observers (e.g., happily surprised versus angrily surprised). This coding scheme was derived from participants spanning multiple ethnicities and races (e.g., Caucasian, Asian, African American and Hispanic). With the inclusion of *compound emotions* that closely mimicked real-life emotional displays, this coding scheme allowed expanding the repertoire of inferred emotions. In the final analysis, I included 20/21 emotions (see Table 1). Prototypical AUs present in $\geq 60\%$ subjects were considered. Appalled and hatred were merged into one category because of similar prototypical AUs corresponding to both. Expressive exemplars are shown in Figure 3.

In summary, I coded 28 *unique emotion categories* on a frame-by-frame basis using distinctive AU signatures, as specified in Table 1. This coding was done programmatically by testing for the *simultaneous occurrence* of these activated AUs—AUs that I had automatically detected from *phase one*. This naturally resulted in multiple emotions being inferred in each frame. Such a culturally-generalizable generalizable approach for emotion inference foregrounding coding of non-prototypical and compound emotions that occur in differentially scaffolded problem-solving contexts can be considered a strength of this article.

*Computing emotion incidence and correlating to learning outcomes (phase three)*

In the *third phase*, I used annotations of emotion inferred on a frame-by-frame basis (from *phase two*) to (a) descriptively compare and contrast the *frequency* with which these emotional states occurred across conditions,
and (b) correlate this frequency to the problem-solving phase and posttest performance (using Spearman’s ρ). Based on prior PS-I work (Loibl et al., 2017), I used two measures of problem-solving phase performance—solution quantity and solution quality. Solution quantity was computed by simply counting the number of unique solutions students developed. Solution quality for the preparatory task was computed by looking at how many rank-ordered pairs were correctly identified (scores from 1 to 6). Students also reported confidence in their generated solution using a 5-point slider.

For posttest scores, I used varimax-rotated principal component analysis to reduce the correlated (binary) posttest scores to a smaller set of important independent composite scores. The rotated factor loadings of three-component principal component analysis (eigenvalues 1.44, 1.04, and 0.85), which accounted for 83% of the total variance, were in line with the intended differentiation between non-isomorphic conceptual understanding (Items 2 and 3), transfer (Item 4) and isomorphic conceptual understanding.

**Figure 3.** Images exemplifying discrete emotion categories from (Du et al., 2014) database. Categories are (A) neutral, (B) happy, (C) sad, (D) fearful, (E) angry, (F) surprised, (G) disgusted, (H) happily surprised, (I) happily disgusted, (J) sadly fearful, (K) sadly angry, (L) sadly surprised, (M) sadly disgusted, (N) fearfully angry, (O) fearfully surprised, (P) fearfully disgusted, (Q) angrily surprised, (R) angrily disgusted, (S) disgustedly surprised, (T) appalled, (U) hatred, and (V) awed. Reprinted from “Compound facial expressions of emotion,” by (Du et al., 2014), Proceedings of the National Academy of Sciences, 111(15), E1454-E1462. Copyright 2014 by National Academy of Sciences. Reprinted with permission.
(Item 1). I used estimates of the derived component scores as representative of these three posttest dimensions. See supplementary materials for the problem design rationale for posttest items.

**Discovering emotion dynamics (phase four)**

In the fourth phase, I discovered frequently co-occurring and temporally contingent patterns of emotions using a closed sequential pattern mining algorithm with time-constraints (Fournier-Viger et al., 2016). Technically, for sequences \( s_1 = \langle t_1, X_1 \rangle, \langle t_2, X_2 \rangle, \ldots, \langle t_n, X_n \rangle, s_2, \ldots, s_n \) where each itemset \( X_x \) is annotated with a timestamp \( t_{x0} \), the algorithm works by extracting all patterns that appear in at least \( S \% \) sequences, that is exceed a minimum support level \( S \). Closed patterns are a compact representation of all sequential patterns, that is, they are not strictly included in another pattern having the same support. For the present analysis, this means that each students’ inferred emotions from phase two (along with the corresponding timestamp) comprised one sequence (e.g., \( s_1 = \langle t_1, \text{surprise} \rangle, \langle t_2, \text{surprise}, \text{interest} \rangle, \ldots, \langle t_n, \text{anger} \rangle \)). Each frame within this sequence comprised one itemset (e.g., \( \langle t_1, \text{surprise} \rangle \)). The timestamp for the first itemset \( (t_1) \) and last itemset \( (t_n) \) corresponded to the beginning and end of the problem-solving and/or instruction phases for a student. Mining closed sequential patterns allowed filtering uninteresting patterns, without losing information. As an index of generalizable empirical evidence, I used 75\% for minimum support (minimum percentage of sequences containing a sequential pattern). Further, 15 frames (0.5 seconds) and 45 frames (1.5 seconds) were the minimum and maximum time interval allowed between itemsets of a sequential pattern (both between two successive itemsets, and between the first itemset and the last itemset). See supplementary materials for the underlying methodological rationale.

**Results**

As previously described, the first problem-solving attempt and follow-up instruction phases were identical for all experimental conditions. Therefore, here I focus on results from the second problem-solving attempt where different kinds of scaffolding were instantiated.

**Emotion incidence (RQ1)**

Based on the aforementioned analysis plan, I first looked at the difference in the incidence of emotions across the conditions. Figure 4–6 depict the percentage of video frames coded with emotion categories based on the three coding schemes. In summary, the two key takeaways from the results of RQ1 are that (a) encountering new information from the scaffolds in the
Figure 4. Violin plots depicting the percentage of video frames coded with different emotion categories (Cordaro et al., 2018 coding scheme).
Figure 5. Violin plots depicting the percentage of video frames coded with different emotion categories (Keltner et al., 2019 coding scheme).
Figure 6. Violin plots depicting the percentage of video frames coded with different emotion categories (Du et al., 2014 coding scheme).
failure-driven and success-driven conditions led to greater observable displays of surprise and interest relative to the Productive Failure condition, where simply students kept generating more solutions, and that (b) the incidence of confusion, a signature of cognitive disequilibrium, was plausibly the lowest in the success-driven condition, where students were nudged toward the canonical solution via structuring scaffolds. Note that because the study samples were relatively small, I focused attention on Cohen’s $d$ measure of effect size to evaluate differences in emotion incidence across the experimental conditions rather than (non)-significance of results obtained after accounting for multiple comparisons. Several observations are worth pointing out.

**Knowledge emotions**

Across all three coding schemes, the two scaffolding conditions had a higher incidence of surprise, compared to the Productive Failure condition. Trends for the failure-driven condition (Cohen’s $d = 0.32, 0.34, 0.18$) were relatively stronger than those for the success-driven condition (Cohen’s $d = 0.21, 0.13, 0.17$). Under the assumptions of a normal distribution, these effect sizes represent up to a 59% likelihood that a random student picked from the two scaffolding conditions will have a higher incidence of surprise than someone from the Productive Failure condition. The incidence of interest followed a similar pattern as surprise, however with weak effect sizes (Cohen’s $d < |0.2|$). The incidence of confusion, when assessed using Cordaro’s and Keltner’s scheme, was however higher in the Productive Failure condition, compared to the failure-driven (Cohen’s $d = 0.24, 0.37$) and success-driven (Cohen’s $d = 0.43, 0.43$) conditions. This suggests up to a 62% likelihood of a relatively lower incidence of confusion in a random student from the success-driven condition. These results partially support hypothesis 1a, which predicted the success-driven condition to have the least frequency of knowledge emotions, however only for the relative incidence of confusion.

**Hostile emotions**

Similar levels of anger, disgust, and contempt were observed across all conditions (Cohen’s $d < |0.2|$). Compound emotion categories involving subtle variations of surprise (sadly surprised, angrily surprised) and anger (sadly angry, fearfully angry, appalled/hatred) also had relatively high and similar incidence across all conditions. Thus, hypothesis 1b, which predicted a higher incidence of hostile emotions in the failure-driven and Productive Failure conditions, could not be supported.
**Self-conscious emotions**

The failure-driven (Cohen’s $d = 0.32$) and Productive Failure (Cohen’s $d = 0.41$) conditions had a higher incidence of shame, compared to the success-driven condition. This reflects up to a 61% likelihood that a random student picked from the failure-driven and Productive Failure conditions will have a higher incidence of shame than someone from the success-driven condition. These results support hypothesis 1c, which predicted the incidence of self-conscious emotions to be the lowest in the success-driven condition.

**Pleasurable emotions**

Finally, all experimental conditions had similar levels of happiness (Cohen’s $d < |0.2|$). These results do not support hypothesis 1d, which predicted students in the success-driven condition to have the highest occurrence of pleasurable emotions.

**Correlation between emotion incidence and learning outcomes (RQ2)**

After investigating the difference in the incidence of emotions across the experimental conditions, I further evaluated their correlation with learning outcomes. In summary, the two key takeaways from the results of RQ2 are that (a) in the failure-driven and Productive Failure conditions, observable displays of anger, disgust and hatred were significantly positively correlated with scores on the non-isomorphic conceptual understanding and/or transfer posttest, and (b) observable displays of other negative (hostile) emotions like contempt and pain and even positive emotions like happiness were in fact significantly negatively correlated to posttest scores. I describe these and remainder of the results below.

**Productive failure condition**

For the Productive Failure condition, none of the inferred emotions were significantly correlated to problem-solving phase performance. However, when evaluating correlations with posttest performance, except for anger ($\rho = 0.38^*$), all other significant correlations with non-isomorphic and/or transfer assessment scores were negative—contempt ($\rho = -0.36^*$), pain ($\rho = -0.39^*$), happily surprised ($\rho = -0.36^*$), disgust ($\rho = -0.41^*$) and its variations—sadly disgusted ($\rho = -0.35^*$), angrily disgusted ($\rho = -0.38^*$).

Taken together, these results for the Productive Failure condition (a) do not support hypothesis 2a, which predicted a positive correlation between the incidence of self-conscious emotions and learning outcomes, (b) partially support hypothesis 2b, which predicted differences in the correlations between hostile emotion categories and learning outcomes, however only for anger and contempt, (c) do not support hypothesis 2c, which predicted the incidence of knowledge emotions to be positively correlated with
learning outcomes, and finally (d) partially support hypothesis 2d, which predicted a negative correlation between the incidence of pleasurable emotions and learning, however only for a compound variation of happiness (that is, happily surprised).

**Success-driven condition**
For the success-driven condition, problem-solving phase performance, as measured via solution quantity had a positive correlation to emotion incidence for the categories of happiness (\(\rho = 0.34^*\)), angrily disgusted (\(\rho = 0.38^*\)), awe (\(\rho = 0.38^*\)), and three variations of surprise—happily surprised (\(\rho = 0.40^*\)), fearfully surprised (\(\rho = 0.38^*\)), disgustedly surprised (\(\rho = 0.36^*\)). Also, solution quality was positively correlated to the incidence of pride (\(\rho = 0.40^*\)). On the other hand, when evaluating correlations with posttest performance, only embarrassment (\(\rho = −0.40^*\)) and pain (\(\rho = 0.35^*\)) were significantly correlated with non-isomorphic outcomes. Correlations between the incidence of other emotions and posttest outcomes were weak and/or non-significant, suggesting a failure to reject the null hypothesis of no monotonic relationship between these variables.

Taken together, these results for the success-driven condition (a) support hypothesis 2a, which predicted no relationship between the incidence of self-conscious emotions and learning outcomes, (b) support hypothesis 2b, which predicted no relationship between the incidence of hostile emotions and learning outcomes, (c) partially support hypothesis 2c, which predicted the incidence of knowledge emotions to be positively correlated with learning outcomes, however only for compound variations of surprise (that is, happily surprised, fearfully surprised, disgustedly surprised), and finally (d) partially support hypothesis 2d, which predicted a positive correlation between the incidence of pleasurable emotions and learning outcomes.

**Failure-driven condition**
Finally, for the failure-driven condition, evaluating correlations with the problem-solving phase performance suggested that only surprise (as assessed by Keltner’s scheme), was positively correlated to solution quantity (\(\rho = 0.44^{**}\)). Solution quantity showed negative correlations with interest (\(\rho = −0.48^{**}\)) and pride (\(\rho = −0.40^*\)), while solution quality showed negative correlations with interest (\(\rho = −0.36^*\)) and when students were inferred to be in a happily disgusted (\(\rho = −0.32^*\)) state. Correlations between the incidence of other emotions and problem-solving phase performance were weak and/or non-significant, suggesting a failure to reject the null hypothesis of no monotonic relationship between these variables. Notably though, despite the non-significant (although moderately positive) correlation between the incidence of shame and students’ solution quality (\(\rho = 0.11\ n.s\)) and confidence in the generated solution (\(\rho = 0.11\ n.s\)), a moderation analysis
revealed that these effects were more pronounced for students with high self-reported domain-specific self-esteem. A multiple regression model with the incidence of shame, domain-specific self-esteem (assessed via the questionnaire from Harter (2012)), and their interaction term, revealed a marginally significant and positive interaction effect ($\beta = 0.06, p = .05$). High self-esteem triggers a positive attributional style toward success and failure, that is, positive events to stable, global, and internal causes, and negative events to temporary, specific, or external causes. On the other hand, students with low self-esteem are predisposed to take failures personally and attribute successes to external causes (Fielstein et al., 1985).

On the other hand, when evaluating correlations with posttest performance, the incidence of disgust was positively correlated to scores on the transfer ($\rho = 0.41^*$) and non-isomorphic conceptual understanding ($\rho = 0.47^*$) assessment. The incidence of contempt ($\rho = -0.43^*$) and happiness ($\rho = -0.47^*$), on the contrary, were negatively correlated to transfer assessment scores. Finally, a higher incidence of appalled/hatred was positively correlated to non-isomorphic ($\rho = 0.42^*$) but negatively correlated to isomorphic conceptual understanding scores ($\rho = -0.44^*$).

Taken together, these results for the failure-driven condition (a) partially support hypothesis 2a, which predicted a positive correlation between the incidence of self-conscious emotions and learning outcomes, (b) support hypothesis 2b, which predicted differences in the correlations between hostile emotion categories and learning outcomes, (c) partially support hypothesis 2c, which predicted the incidence of knowledge emotions to be positively correlated with learning outcomes, however only for surprise, and finally (d) support hypothesis 2d, which predicted a negative correlation between the incidence of pleasurable emotions and learning outcomes.

**Emotion dynamics (RQ3)**

Finally, I looked at the difference in emotion dynamics across the experimental conditions. Table 2 synthesizes the sequential patterns of emotions that were discovered for each experimental condition, organized by the three coding schemes I used. In summary, the three key takeaways from the results of RQ3 are that (a) the number and diversity of discovered sequential patterns of emotions in the failure-driven condition were higher than the Productive Failure condition, and nearly two times those discovered in the success-driven condition, (b) the failure-driven condition had exclusive sequential patterns comprising shame, and (c) all experimental conditions had dynamics comprising knowledge emotions associated with thinking and comprehending (e.g., surprise), hostile emotions signaling negative evaluation of the task situation (e.g., anger), and pleasurable emotions (e.g., happiness). Representative visual examples from my study are presented in
## Table 2. Dynamics of inferred emotions for the second attempt during the problem-solving phase (post-scaffold).

| Condition       | Cordaro et al., 2018                                      | Keltner et al., 2019                                      | Du et al., 2014                                      |
|-----------------|-----------------------------------------------------------|-----------------------------------------------------------|-----------------------------------------------------|
| Success-driven  | Anger → Anger → Anger                                      | Surprise → Surprise → Surprise                            | Angrily surprised → Angrily surprised                |
|                 | Contempt → Contempt                                        | Interest → Interest                                        | Sadly angry → Sadly angry → Sadly angry              |
|                 | Surprise → Surprise                                        |                                                           | Disgust → Appalled/hatred                             |
|                 | Anger, Contempt → Anger                                    |                                                           | Anger → Sadly fearful, Fearfully angry               |
|                 | Surprise → Anger                                            |                                                           | Happy, Happily disgusted → Happy                     |
| Productive Failure | Anger → Anger → Anger                                      | Surprise → Surprise → Surprise                            | Sadly angry → Sadly angry                            |
|                 | Contempt → Contempt                                        | Interest → Interest                                        | Disgust → Disgust, Appalled/hatred                   |
|                 | Surprise → Surprise                                        | Disgust → Disgust                                         | Sadly fearful, Fearfully angry → Angrily surprised   |
|                 | Contempt → Contempt → Anger                                 |                                                           | Anger → Sadly disgusted                              |
|                 | Surprise → Anger                                            |                                                           | Happy, Happily disgusted → Happy                     |
| Failure-driven  | Anger → Anger → Anger                                       | Surprise → Surprise                                        | Angrily surprised → Sadly surprised, Angrily surprised|
|                 | Anger → Contempt                                            | Interest → Interest → Interest                             | Surprise → Surprise, Fearfully surprised, Awe        |
|                 | Surprise → Surprise                                        | Disgust → Disgust                                         | Anger → Anger, Sadly disgusted                       |
|                 | Shame → Shame → Shame                                       |                                                           | Anger, Sadly fearful → Fearfully angry               |
|                 | Shame → Anger                                               |                                                           | Happy, Happily disgusted → Happy                     |

A sequence X → Y, Z can be interpreted to mean that the occurrence of an inferred emotion X at time ‘t’ is followed by simultaneous occurrence of inferred emotions Y and Z at time ‘t + [0.5–1.5] seconds.’
Note that unlike exemplar images from the work of Du et al. (2014) in Figure 3, facial movements characterizing different emotion dynamics during the preparatory problem-solving task are much more subtle—the corresponding AUs are detected as activated, but the intensity of activation is not necessarily very high. I elaborate more below.

**Intensity of emotional experience**

During the second problem-solving attempt, I discovered a greater number (and diversity) of sequential patterns of emotion \( n = 12,927 \) in the failure-driven condition, relative to the Productive Failure \( n = 11,183 \) and success-driven \( n = 6684 \) conditions. This result supports hypothesis 3a, which predicted a wider variability in emotion dynamics for the failure-driven and Productive Failure conditions.

**Exclusive dynamics in the failure-driven condition**

Additionally, I found exclusively occurring sequences comprising shame when failure-driven scaffolds were introduced during the second problem-solving attempt. Despite a similar percentage of video frames being inferred as shame \( (\text{Cohen’s } d < |0.2|) \) in both the failure-driven and Productive Failure conditions (similar incidence), none of the frequently occurring sequential patterns in the Productive Failure condition comprised shame (different dynamics). A similar trend was seen for the success-driven condition. In a similar vein, despite the relatively higher incidence of confusion in the Productive Failure condition, none of the frequent sequential patterns comprised confusion.
Dynamics across all experimental conditions

A frequently occurring set of sequential patterns across all conditions comprised surprise (existence of an expectation with which the stimulus disagrees), along with its variations like angrily surprised, sadly surprised and fearfully surprised, and sometimes followed by emotional displays of awe. Across all conditions, I also found sequential patterns comprising interest—the incidence of interest in sequential patterns from the failure-driven condition \((n = 254)\), however, was nearly four times compared to the success-driven \((n = 58)\) and Productive Failure \((n = 62)\) conditions. As evident from Table 2, across all conditions, I further discovered sequential patterns comprising several temporally contingent and co-occurring negative emotions like anger, contempt, and disgust. A look at compound emotions, coded via the scheme by Du et al. (2014), revealed sequences where students were inferred to display subtle variations of basic negative emotions such as sadly angry, sadly fearful, fearfully angry, sadly disgusted, and happily disgusted. Finally, I discovered contingent patterns of happiness across all conditions.

Discussion

The goals of the presented process-focused analysis were to expand the explanatory basis of PS-I by (a) discovering moment-by-moment determinants of students’ affective states during preparatory problem-solving and follow-up instruction, (b) understanding how the incidence and temporal dynamics of these states vary based on manipulating the problem-solving context with scaffolding strategies (failure-driven, success-driven, none), and (c) assessing the extent which affective states might explain learning. Note that the emotions I measured in the current study are state emotions that occur in the moment and are closely influenced by particular situations within the task context. In contrast, trait emotions reflect general affective dispositions to respond. For instance, interest in STEM disciplines that is usually assessed via questionnaire ratings would fall into the latter formulation of emotion/affect as a trait. Because state and trait emotions are distinct (although mutually influencing) categories, the presented results do not dispute previously established work on the criticality of affective dispositions for learning in STEM (Maltese & Tai, 2011). Notably, however, they do provide additional empirical evidence for a more diverse palette of emotions, especially negative emotions, that might be useful to consider when researching students’ affective dispositions, especially in preparatory problem-solving contexts.

I synthesize explanations for five key results of the work below, and argue for the meaningfulness of these results by (a) interpreting them in the context of design features of the PS-I learning design (e.g., nature of the problem,
(1) Preparatory problem-solving, especially in the presence of failure-driven scaffolding is an intense emotional experience, indexed by the differences across the discovered sequential patterns of emotions.

(2) Students exposed to failure-driven scaffolding in PS-I show exclusive dynamics comprising emotional displays of shame, a self-conscious emotion associated with both metacognitive and cognitive benefits.

(3) Knowledge emotions are widely prevalent in PS-I and the presence of failure-driven scaffolding creates opportunities for relatively greater emotional displays of surprise, interest, and confusion in preparatory problem-solving.

(4) Hostile emotion categories differentially impact learning in PS-I, with the incidence of anger and disgust showing positive associations and the incidence of contempt showing a negative association.

(5) Pleasurable emotions (e.g., happiness) in PS-I positively associate with isomorphic posttest outcomes but negatively associate with non-isomorphic and/or transfer posttests.

**Number of sequential patterns**

The relatively greater number of frequently occurring sequential patterns of emotion in the failure-driven condition can be explained by the people’s general sensitivity and greater response likelihood to negative stimuli compared to positive stimuli (Carretié et al., 2001). Alongside these process measures, an analysis of aggregated measures of affect from retrospectively reported problem-solving experiences (Sinha & Kapur, 2021a) corroborates these results. I found that students in the failure-driven condition had higher scores on self-reported negative affect relative to the Productive Failure condition, and higher scores on both positive and negative affect relative to the success-driven condition. The PANAS instrument (Watson et al., 1988) was used to assess perceived affect. Positive affect (10 items, \( \alpha = 0.91 \)) reflects the extent to which one subjectively experiences positive moods such as joy, interest, and alertness. Negative affect (10 items, \( \alpha = 0.89 \)) reflects feelings of emotional distress, defined by the common variance between anxiety, sadness, anger, and other unpleasant emotions. See supplementary materials for sample items.

Stronger emotional experiences during preparatory problem-solving, especially negative emotions, can improve memory accuracy by increasing the ability to access mood-congruent information (Bower, 1981). This has
the potential to give failure-driven condition students a relative advantage over their counterparts, especially when undertaking non-isomorphic and/or unfamiliar transfer posttests that might evoke similar negative emotions.

**Self-conscious emotions during complex problem-solving**

As a self-conscious emotion (Silvia, 2009), experiencing *shame* signals to an individual that they have a sense of self and the ability to reflect upon what the self has done. This reflection serves as an important metacognitive function in re-orienting an individual with their goals, values, and self-image and is likely to be associated with more accurate self-evaluations of problem-solving performance. In the failure-driven condition with frequently occurring sequential patterns of shame, empirical evidence for this conjecture is bolstered by (a) moderately positive correlation between the incidence of shame and students’ expressed confidence for the non-isomorphic conceptual understanding (ρ = 0.212 n.s) and transfer (ρ = 0.203 n.s) posttests, and the fact that (b) these students were in fact metacognitively well-calibrated at the posttest, as reflected, via a non-significant one-sample t-test comparing how the gap between scaled values of confidence judgments and actual performance differs from 0 (see Sinha and Kapur (2021a) for details). In contrast, the correlation between the incidence of shame and students’ confidence at the posttest was very weak for the success-driven (ρ = −8.396e-4 n.s, −0.020 n.s) and Productive Failure (ρ = −0.047 n.s, −0.013 n.s) conditions. Students’ metacognitive judgments both influence and are influenced by their awareness of the effectiveness of applied problem-solving strategies. Experiencing shame in such situations might serve to mitigate students’ reliance on biased evaluations of problem-solving performance and prevent, for instance, premature termination of solution strategy exploration.

When originally conceptualized as an appraisal of the self as fundamentally flawed, shame was seen as aversive, unpleasant, and a trigger for orienting people to avoid failure and its consequences (Lewis, 1971). However, a radically different view has been proposed in more recent developments, where the intense dysphoria of shame is perceived to be a motivator for sensemaking, especially if such failures appear more reparable. In a PS-I setting, the learning design provides the opportunity for students to address the cause or consequences of one’s failure because they can re-do or re-attempt the task, get situational (e.g., syntax-level) feedback from the problem-solving environment, and ask for more scaffolds if necessary. Meta-analytic evidence suggests that shame has a positive link to such a constructive approach when performance failures such as doing poorly on an intellectual task are more reparable (Leach & Cidam, 2015). The more pronounced correlational effects between the incidence of shame and problem-solving
phase performance in the failure-driven condition, especially for those students with high self-reported domain-specific self-esteem, provide further empirical support for this evidence. Plausibly, such students are more likely to be motivated to engage in sensemaking because of the dysphoria of shame. No such interaction effect existed for the Productive Failure and success-driven conditions. Taken together, these results call for a reconception of shame as a positively influential emotional state within a preparatory problem-solving context, especially when failure-likelihood is high.

**Knowledge emotions during complex problem-solving**

*Surprise, interest, and confusion*, which are often discussed under knowledge emotions associated with thinking and comprehending (Silvia, 2009), motivate exploratory action.

**Surprise**
The generative nature of preparatory problem-solving tasks makes *surprise* a natural candidate to trigger a sense of discovery, a sense of making way through learning material that suddenly impresses students as important or significant. More generally, surprise, as assessed by physiological measures such as pupillary response, has been shown to enhance learning and boost memory when expectancy-violating information is encountered during prediction tasks (Brod et al., 2018). Although being open to surprise may be demanding, often unexpected patterns in the presented datasets can serve as novel and insightful moments to trigger sensemaking, for example, starkly contrasting scatterplots despite similar correlations. In contrast to the Productive Failure condition where students simply keep generating more solutions to the task, encountering new information from scaffolds presented in the failure-driven and success-driven conditions and gaining previously unknown insights about the data and/or the targeted phenomena is likely to aggravate the experience of surprise. Indeed, empirical evidence from the success-driven condition suggests that the incidence of *happily surprised, fearfully surprised, and disgustedly surprised* was positively correlated to students generating more solutions during the problem-solving phase—solutions, which stem from scaffolds nudging students toward more manageable directions.

**Interest**
The incidence and dynamics of *interest* signal that students not only appraise the problem-solving task as novel and unexpected but also view it as comprehensible, that is exhibit a high coping-potential appraisal. Students with a high interest in the learning materials are more likely to process the presented information deeply and persist longer at the learning task (Harackiewicz et al., 2016). This might motivate the preparation for learning
from instruction (Lammina & Chase, 2019). An analysis of the aggregated measure of state curiosity from retrospectively reported problem-solving experiences (Sinha & Kapur, 2021a) corroborates these results. I found that students in the failure-driven and success-driven conditions had higher scores on self-reported state curiosity relative to the Productive Failure condition. State curiosity (9 items, $\alpha = 0.86$), which reflects students’ desire to know more about the canonical answer to fill their knowledge gaps, was assessed using the state-trait curiosity inventory (Naylor, 1981). See supplementary materials for sample items.

**Confusion**

Unlike interest, the incidence of confusion involves appraising the problem-solving situation as hard to understand. Results suggested that students in the success-driven condition experienced the least confusion, while both failure-driven and Productive Failure conditions had a relatively higher incidence of confusion. This is plausible because students are more likely to encounter impasses when failure is explicitly or implicitly induced. By creating a perceived vacuum, an impasse can help see the information needs and knowledge gaps to be filled more clearly, thereby leading to better focus on the relevant domain principles in a subsequent learning phase (Glogger-Frey et al., 2015). Furthermore, the confusion caused by an impasse initiates inquiry, and engaging in such forms of inquiry has been empirically found to promote learning gains as areas of confusion are overcome (D’Mello et al., 2014).

**Hostile emotions during complex problem-solving**

Anger, disgust, and contempt, as a hostility triad of emotions that are frequently experienced together (Izard, 1977; Silvia, 2009), all involve disapproval or negative evaluation of the current situation an individual is confronted with.

**Anger**

Research on the role of perception of potential threats has found that anger is often triggered by viewing a situation as personally relevant and inconsistent with what one is trying to achieve, especially when such a situation is deliberately designed by others (Roseman, 1991). As a reaction to respond to goal blockage that could be caused here by attending to a problem-solving situation with non-readily verifiable outcomes, anger encourages reflection and leads to motivation to act and restore equilibrium by marshaling needed resources (Harmon-Jones et al., 2014).

Design features of my PS-I environment, for example, forcing students to explain their reasoning in words and rank-order datasets based on criterion, are further likely to engage students in confrontational response strategies.
Qualitative evidence for this claim comes from observations of synchronous screen recordings, where I often found episodes of highly intense student action. There was frequent switching between seeking help from inbuilt programming libraries to write executable programming code, and interpreting resulting patterns in the presented datasets to formulate an appropriate (non-)mathematical response to the rank-ordering task. Positive correlations of anger with non-isomorphic and/or transfer posttest scores in the Productive Failure condition also provide empirical support for this conjecture. Incidence of fearfully angry might suggest that students’ sense of self-preservation stemming from fear helps them navigate unexpected obstacles, fueled by moderate levels of anger.

**Disgust**

Disgust, on the other hand, might be triggered by problem-solving situations where students are likely to feel betrayed or when they perceive the situation to be hypocritical (Miller, 1998). One potential source for the incidence of disgust in my study setting might be the problem design that uses variant-invariant features, for example, keeping certain features similar across contrasting datasets (e.g., same descriptive statistics) but systematically varying other features (e.g., visualizations). Here, presenting students with only descriptive statistics and asking them to rank-order datasets based on a criterion is likely to induce disgust—a strict ordering cannot be derived if problem-features are similar, and an integrated understanding of these and other potentially unknown problem features is needed for progress.

Additionally, scaffolds in the failure-driven condition, which increase the degrees of freedom during problem-solving, contrast the usual help-seeking perception of decreasing uncertainty and making problem-solving more tractable. This might further aggravate overt emotional displays of disgust. However, despite experiencing disgust, a deeper exploration of the problem-space that would not have been possible in the absence of deliberately designed failure-driven scaffolds, could explain why the incidence of disgust was positively correlated to transfer posttest scores. Incidence of variations of disgust like sadly disgusted and angrily disgusted in the Productive Failure condition, however, might reflect dissatisfaction with problem-solving efforts due to its deliberately designed ill-structured nature, despite students holding a general passion to actively participate. Remember that students self-selected to participate after passing a pre-screening programming quiz, that is, scoring ≥7/10. This might explain why the incidence of these compound emotional states was found to be negatively correlated with non-isomorphic and/or transfer posttest scores.

Finally, it can be posited that as students keep generating more solutions during their second problem-solving attempt, mixed emotional reactions are likely to be evoked. On one hand, there might be a sense of relief
because of already having created a sufficiently high number of solutions. On the other hand, there might also be dissatisfaction due to the general awareness that high solution generation need not necessarily correlate with chances of stumbling onto the canonical solution, especially when the targeted learning concept is yet to be learned. This conflict might result in a higher number of overt emotional displays of students being in a happily disgusted state. Explicit failure-driven scaffolds are expected to further strengthen this conflict—this might explain why the incidence of happily disgusted for students in the failure-driven condition was negatively correlated to solution quality in the problem-solving phase. An analysis of the aggregated measure of cognitive dissonance from retrospectively reported problem-solving experiences (Sinha & Kapur, 2021a) corroborates these results. I found that students in both the Productive Failure and failure-driven conditions had higher scores on self-reported cognitive dissonance relative to the success-driven condition. Cognitive dissonance (6 items, $\alpha = 0.53$), which reflects a state of discomfort associated with the detection of conflicting concepts, was assessed using the scale proposed by Levin et al. (2013). See supplementary materials for sample items.

**Contempt**

As opposed to anger that is confrontational, contempt reflects a view of disapproval and indifference toward the problem-solving situation (Fischer & Giner-Sorolla, 2016). As students keep generating multiple solutions and representation methods during their pre-scaffold and post-scaffold attempts, lack of concrete accuracy feedback might result in them judging the situation as unworthy of their sustained efforts/attention, resulting in overt facial movements reflective of contempt. This view was endorsed in retrospective accounts of students’ response strategy to scaffolds presented during the second problem-solving attempt, which I assessed via a single item adapted from Chinn and Brewer (1993). Empirical evidence showed that 46.3% of students’ response strategies fell into the categories of ignoring/rejecting evidence from the scaffold, excluding the scaffold information from one’s knowledge domain, dealing with the scaffold later, and reinterpreting evidence from the scaffold to be congruent with one’s original reasoning. And indeed, as expected, the incidence of contempt, which might be reflective of such a dismissive nature toward the scaffolds, was negatively correlated with non-isomorphic and/or transfer posttest scores for both scaffolding conditions.
Pleasurable emotions during complex problem-solving

In contrast to interest that motivates trying new problem-solving and reasoning strategies, happiness is linked to attachments to learning strategies that have proved rewarding in the past (Silvia, 2008). Working on the preparatory task across two attempts (40–45 minutes) should give students enough time to revel in their happiness, as and when they get positive feedback on a task action and/or when solution pathways start becoming clearer. The likelihood of such events is higher for students in the success-driven condition, who are nudged toward the canonical answer—positive correlations between the incidence of happiness and the amount of solution generation for these students lends support to this explanation.

Conclusion

To conclude, I focused on nonverbal behaviors of the face, operationalized as facial landmarks, to infer naturally-occurring emotions and their sequential dynamics in scaffolded problem-solving. I attempted to overcome some criticisms of the discrete conceptualization of emotion and the common view framework for emotion inference by inferring additional non-prototypical and compound emotions, using multiple culturally-generalizable coding schemes for emotion inference, and manipulating the context in which facial movements occur using failure-driven and success-driven scaffolding. Results showed not just a greater number and diversity of emotion dynamics, but also exclusive dynamics involving shame for the failure-driven condition. By correlating the incidence of different emotion categories with the problem-solving phase and posttest performance, I demonstrated how the emotional roller coaster ride students experience in preparatory problem-solving might affect learning.

Note that because I did not manipulate the incidence of negative emotions, the implications of my results are not to incite negative emotions and/or expose students to learning conditions where they dwell on the discovered negative emotions for extended periods. Indeed, this can result in a negative thinking spiral that can deplete the cognitive ability to problem-solve proactively. However, what my results suggest is that overt changes in facial movements reflective of students potentially experiencing negative emotional states, can in fact be beneficial. Experiencing moderate levels of negative emotions keeps one alerted of challenges requiring more focused attention, and assists in comprehending conflicting information (Ivtzan et al., 2015; Kashdan & Biswas-Diener, 2014).

Recent meta-analytic evidence (MacCann et al., 2020) suggests that knowledge about the causes and consequences of emotions, and knowing how to manage emotional situations are core aspects of emotional intelligence.
that predict academic performance. Condemning emotions may therefore obstruct practicing emotional intelligence. By using the empirical evidence about the incidence of particular categories of emotions and their frequently occurring dynamics in a preparatory problem-solving context, teachers can devote resources that focus on helping students understand and manage these emotions (e.g., Brackett et al.’s (2019) RULER approach). When students broaden their emotional vocabulary or describe their emotions more precisely, they can take the first step toward accurately diagnosing emotional states and choosing an appropriate response strategy. Such an investment is likely to reap benefits for both achievement and their emotional well-being, in particular when focused on emotions that I found to be positively correlated with learning outcomes. On the other hand, when emotional influences may be unwanted such as the ones I found to be negatively correlated to learning outcomes, and/or in cases of extreme disengagement such as when a student completely withdraws from the task, strategies to reduce the effects of such emotions on task progress may be initiated (Lerner et al., 2015). Replication of these results including more longitudinal work would be needed before generalizing these claims.

An intrinsic limitation of the nature of this work stems from a prevailing tension (Reiter-Palmon et al., 2017) in multimodal learning analytics. How to preserve enough of the complexity of inferred behaviors in a theory-driven analysis framework to obtain valid scientific insight, while simultaneously attain computational feasibility for automatizing inference? For instance, here I (a) used simple binary features reflecting whether or not a particular facial action unit was activated in a given frame, and (b) preserved facial movement data at its original time granularity to maximize the information available for all further downstream analyses. Future work might incorporate the intensity of activation of facial action units as well evaluate the effectiveness of varying the time scale for multimodal fusion of facial action units. Further, despite using facial action unit recognition algorithms that were trained on a diverse participant demographic varying by age, gender, ethnicity, etc. (Baltrusaitis et al., 2018), I acknowledge that the inherent bias stemming from using these algorithms for automated analyses can never be completely eliminated. More learning research is also needed to triangulate empirical support regarding whether prototypical facial expressions of emotion are more than merely stereotypes or oversimplified beliefs. Finally, it is hard to completely reconstruct reality through video-based observations, despite my attempts toward making the data collection more ecologically valid by (a) performing manipulation checks and administering self-reports targeting specific psychological variables to evaluate intervention effectiveness, (b) allowing time before the beginning of the video recording for the problem-solving phase (e.g., via pretests and incoming profile questionnaires) so that participants settle into normal behaviors, (c) using data
capture devices with low obtrusiveness to target nonverbal behaviors of the face that are relatively difficult to manipulate in contrast to verbal expressions, and (d) reducing participants’ discomfort because of being watched by providing transparency in advantages of participation upfront along with clarifying no risks regarding formal academic evaluation.

As part of the next steps within my research agenda on studying the interplay of cognition and affect within failure-first instructional contexts, I have begun investigating how to best use students’ negative emotional experiences to provide remedial emotional scaffolding that can channel students toward productive exploration. My take is that emotional scaffolding strategies should foreground not just eliciting/managing students’ positive emotional experiences and simply eliminating negative emotions (see Quoidbach et al. (2015) for such a review), but rather adopt a more balanced approach and appropriately acknowledge the tradeoffs of both positive and negative emotions. After all, my results show that negative emotional states, as evidenced by facial movements, within problem-solving scaffolding that is primarily failure-driven, can act as catalysts for learning.

Acknowledgments

I thank Stefan Wehrli and Giordano Giannoccolo (Decision Sciences Lab, ETH Zürich) and Maya Spannagel (University of Zürich) for help with running the study. Data processing was performed using ETH Euler scientific computing cluster. Thanks to Manu Kapur, Elsbeth Stern, Eleni Kyza and Catherine Chase for their critical comments on an initial manuscript iteration that greatly improved this work. I also appreciate the helpful feedback from colleagues at the Professorship for Learning Sciences and Higher Education, ETH Zürich. Special thanks to Carolyn Rosé, Patrick Jermann, Pierre Dillenbourg and Justine Cassell for nudging me towards the Learning Sciences in the formative years of my career and for their ongoing support and encouragement. Finally, I express my sincere thanks to the reviewers and editors of JLS for their invaluable mentorship throughout the review process.

ORCID

Tanmay Sinha http://orcid.org/0000-0003-3069-2899

References

Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L. P. (2018). Openface 2.0: Facial behavior analysis toolkit. In 13th IEEE International Conference on Automatic Face & Gesture Recognition (pp. 59–66). Xi’an, China: IEEE.
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. https://doi.org/10.1177/1529100619832930
Bower, G. H. (1981). Mood and memory. *American Psychologist, 36*(2), 129. https://doi.org/10.1037/0003-066X.36.2.129

Brackett, M. A., Bailey, C. S., Hoffmann, J. D., & Simmons, D. N. (2019). RULER: A theory-driven, systemic approach to social, emotional, and academic learning. *Educational Psychologist, 54*(3), 144–161. https://doi.org/10.1080/00461520.2019.1614447

Brod, G., Hasselhorn, M., & Bunge, S. A. (2018). When generating a prediction boosts learning: The element of surprise. *Learning and Instruction, 55*, 22–31. https://doi.org/10.1016/j.learninstruc.2018.01.013

Cabanac, M. (2002). What is emotion?. *Behavioural Processes, 60*(2), 69–83. https://doi.org/10.1016/S0376-6357(02)00078-5

Cahour, B. (2013). Emotions: Characteristics, emergence and circulation in interactional learning. In M. Baker, J. Andriessen, & S. Jarvela (Eds.), *Affective learning together* (pp. 52–70). Routledge. https://doi.org/10.4324/97802030369684

Carretié, L., Mercado, F., Tapia, M., & Hinojosa, J. A. (2001). Emotion, attention, and the ‘negativity bias,’ studied through event-related potentials. *International Journal of Psychophysiology, 41*(1), 75–85. https://doi.org/10.1016/S0167-8760(00)00195-1

Chi, M. T., Glaser, R., & Farr, M. J. (Eds.). (2014). *The nature of expertise*. Psychology Press. https://doi.org/10.4324/9781315799681

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research, 63*(1), 1–49. https://doi.org/10.3102/00346543063001001

Coles, N. A., Larsen, J. T., & Lench, H. C. (2019). A meta-analysis of the facial feedback literature: Effects of facial feedback on emotional experience are small and variable. *Psychological Bulletin, 145*(6), 610. https://doi.org/10.1037/bul0000194

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. https://doi.org/10.1037/emo0000302

D’Mello, S. (2013). A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology, 105*(4), 1082. https://doi.org/10.1037/a0032674

D’Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction, 22*(2), 145–157. https://doi.org/10.1016/j.learninstruc.2011.10.001

D’Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction, 29*, 153–170. https://doi.org/10.1016/j.learninstruc.2012.05.003

Damasio, A. R. (1994). *Descartes’ error: Emotion, reason and the human brain*. Putnam.

Du, S., Tao, Y., & Martinez, A. M. (2014). Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences, 111*(15), E1454–E1462. https://doi.org/10.1073/pnas.1322355111

Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior, 1*(1), 56–75. https://doi.org/10.1007/BF01115465
Ekman, P., Friesen, W. V., O’sullivan, M., Chan, A., Diacyoanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., Scherer, K., Tomita, M., & Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. Journal of Personality and Social Psychology, 53(4), 712. https://doi.org/10.1037/0022-3514.53.4.712

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), Handbook of affective sciences (pp. 572–595). Oxford University Press.

Festinger, L. (1962). A theory of cognitive dissonance. Stanford University Press.

Fielstein, E., Klein, M. S., Fischer, M., Hanan, C., Koburger, P., Schneider, M. J., & Leitenberg, H. (1985). Self-esteem and causal attributions for success and failure in children. Cognitive Therapy and Research, 9(4), 381–398. http://dx.doi.org/10.1007/BF01173088

Fischer, A., & Giner-Sorolla, R. (2016). Contempt: Derogating others while keeping calm. Emotion Review, 8(4), 346–357. https://doi.org/10.1177/1754073915610439

Fournier-Viger, P., Lin, J. C.-W., Gomariz, A., Gueniche, T., Soltani, A., Deng, Z., & Lam, H. T. (2016). The spmf open-source data mining library version 2. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Riva del Garda, Italy: Springer. https://doi.org/10.1007/978-3-319-46131-1_8

Glogger-Frey, I., Fleischer, C., Grüny, L., Kappich, J., & Renkl, A. (2015). Inventing a solution and studying a worked solution prepare differently for learning from direct instruction. Learning and Instruction, 39, 72–87. https://doi.org/10.1016/j.learninstruc.2015.05.001

Harackiewicz, J. M., Smith, J. L., & Priniski, S. J. (2016). Interest matters: The importance of promoting interest in education. Policy Insights from the Behavioral and Brain Sciences, 3(2), 220–227. https://doi.org/10.1177/2372732216655542

Harmon-Jones, E., Price, T. F., Gable, P. A., & Peterson, C. K. (2014). Approach motivation and its relationship to positive and negative emotions. In M. M. Tugade, M. N. Shiota, & L. D. Kirby (Eds.), Handbook of positive emotions (pp. 103–118). Guilford Press.

Harter, S. (2012). Self-perception profile for adolescents: Manual and questionnaires. University of Denver, Department of Psychology.

Ivtzan, I., Lomas, T., Hefferon, K., & Worth, P. (2015). Second wave positive psychology: Embracing the dark side of life. Routledge. https://doi.org/10.4324/9781315740010

Izard, C. E. (1977). Anger, disgust, and contempt and their relationship to hostility and aggression. In C. Izard (Ed.), Human emotions (pp. 329–354). Springer. https://doi.org/10.1007/978-1-4899-2209-0_13

Jones, S. M., McGarrah, M. W., & Kahn, J. (2019). Social and emotional learning: A principled science of human development in context. Educational Psychologist, 54(3), 129–143. https://doi.org/10.1080/00461520.2019.1625776

Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. Educational Psychologist, 51(2), 289–299. https://doi.org/10.1080/00461520.2016.1155457

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. Journal of the Learning Sciences, 21(1), 45–83. https://doi.org/10.1080/10508406.2011.591717

Kashdan, T., & Biswas-Diener, R. (2014). The upside of your dark side: Why being your whole self—not just your “good” self—drives success and fulfillment. Penguin.
Keltner, D., Sauter, D., Tracy, J., & Cowen, A. (2019). Emotional expression: Advances in basic emotion theory. *Journal of Nonverbal Behavior, 43*(2), 133–160. https://doi.org/10.1007/s10919-019-00293-3

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. https://doi.org/10.1016/j.cedpsych.2019.101785

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. https://doi.org/10.1037/pspa0000037

LePine, J. A., LePine, M. A., & Jackson, C. L. (2004). Challenge and hindrance stress: Relationships with exhaustion, motivation to learn, and learning performance. *Journal of Applied Psychology, 89*(5), 883. https://doi.org/10.1037/0021-9010.89.5.883

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual Review of Psychology, 66*(1), 799–823. https://doi.org/10.1146/annurev-psych-010213-115043

Levin, D. T., Harriott, C., Paul, N. A., Zhang, T., & Adams, J. A. (2013). Cognitive dissonance as a measure of reactions to human-robot interaction. *Journal of Human-Robot Interaction, 2*(3), 3–17. https://doi.org/10.5898/JHRI.2.3 Levin

Lewis, B., H. (1971). Shame and guilt in neurosis. *Psychoanalytic Review, 58*(3), 419–438. https://psycnet.apa.org/record/1972-21079-001

Loderer, K., Pekrun, R., & Lester, J. C. (2018). Beyond cold technology: A systematic review and meta-analysis on emotions in technology-based learning environments. *Learning and Instruction, 70*, 101162. https://doi.org/10.1016/j.learninstruc.2018.08.002

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), 4. https://doi.org/10.1007/s10648-016-9379-x

Loibl, K., & Rummel, N. (2014). Knowing what you don’t know makes failure productive. *Learning and Instruction, 34*, 74–85. https://doi.org/10.1016/j.learninstruc.2014.08.004

MacCann, C., Jiang, Y., Brown, L. E., Double, K. S., Bucich, M., & Minbashian, A. (2020). Emotional intelligence predicts academic performance: A meta-analysis. *Psychological Bulletin, 146*(2), 150. https://doi.org/10.1037/bul0000219

Maltese, A. V., & Tai, R. H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among us students. *Science Education, 95*(5), 877–907. https://doi.org/10.1002/sce.20441

Miller, W. I. (1998). *The anatomy of disgust*. Harvard University Press.

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review, 19*(3), 309–326. https://doi.org/10.1007/s10648-007-9047-2

Naylor, F. D. (1981). A state-trait curiosity inventory. *Australian Psychologist, 16*(2), 172–183. https://doi.org/10.1080/00050068108255893

Noroozi, O., Pijeira-Diaz, H. J., Sobocinski, M., Dindar, M., Järvelä, S., & Kirschner, P. A. (2020). Multimodal data indicators for capturing cognitive, motivational, and emotional learning processes: A systematic literature review. *Education and Information Technologies, 25*(6), 5499–5547. https://doi.org/10.1007/s10639-020-10229-w
Pekrun, R., & Perry, R. P. (2014). Control-value theory of achievement emotions. In P. A. Alexander, R. Pekrun, & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 130–151). Routledge. https://doi.org/10.4324/9780203148211-12

Putwain, D. W., Symes, W., & Remedios, R. (2016). The impact of fear appeals on subjective-task value and academic self-efficacy: The role of appraisal. *Learning and Individual Differences, 51*, 307–313. https://doi.org/10.1016/j.lindif.2016.08.042

Quoidbach, J., Mikolajczak, M., & Gross, J. J. (2015). Positive interventions: An emotion regulation perspective. *Psychological Bulletin, 141*(3), 655. https://doi.org/10.1037/a0038648

Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences, 13*(3), 273–304. http://dx.doi.org/10.1207/s15327809jls1303_2

Reiter-Palmon, R., Sinha, T., Gevers, J., Odobez, J. M., & Volpe, G. (2017). Theories and models of teams and groups. *Small Group Research, 48*(5), 544–567. https://doi.org/10.1177/1046496417722841

Roseman, I. J. (1991). Appraisal determinants of discrete emotions. *Cognition & Emotion, 5*(3), 161–200. https://doi.org/10.1080/02699939108411034

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161. https://doi.org/10.1037/h0077714

Schoenfeld, A. H. (1988). When good teaching leads to bad results: The disasters of ‘well-taught’ mathematics courses. *Educational Psychologist, 23*(2), 145–166. https://doi.org/10.1207/s15326985ep2302_5

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16*(4), 475–5223. https://doi.org/10.1207/s1532690xci1604_4

Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher, 27*(2), 4–13. https://doi.org/10.3102/0013189X027002004

Silvia, P. J. (2008). Interest—the curious emotion. *Current Directions in Psychological Science, 17*(1), 57–60. https://doi.org/10.1111/j.1467-8721.2008.00548.x

Silvia, P. J. (2009). Looking past pleasure: Anger, confusion, disgust, pride, surprise, and other unusual aesthetic emotions. *Psychology of Aesthetics, Creativity, and the Arts, 3*(1), 48. https://doi.org/10.1037/a0014632

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. https://doi.org/10.1016/j.learninstruc.2021.101488

Sinha, T., & Kapur, M. (2021b). When problem-solving followed by instruction works: Evidence for productive failure. *Review of Educational Research, 0034654321101919*. Advance online publication. https://doi.org/10.3102/00346543211019105

Sinha, T., & Kapur, M. (2021c). From problem-solving to sensemaking: A comparative meta-analysis of preparatory approaches for future learning. *EdArXiv*. https://doi.org/10.35542/osf.io/83p7e

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. https://doi.org/10.1037/edu0000483
Sweller, J. (2011). Cognitive load theory. In J. P. Mestre & B. H. Ross (Eds.), *Psychology of learning and motivation* (Vol. 55, pp. 37–76). Academic Press. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Tulis, M., & Ainley, M. (2011). Interest, enjoyment and pride after failure experiences? predictors of students’ state-emotions after success and failure during learning in mathematics. *Educational Psychology, 31*(7), 779–807. https://doi.org/10.1080/01443410.2011.608524

Vahey, P., Finzer, W., Yarnall, L., & Schank, P. (2017). CIRCL primer: Data science education. In *CIRCL Primer Series*. USA. http://circlcenter.org/data-science-education

Vosniadou, S. (Ed.). (2009). *International handbook of research on conceptual change*. Routledge. http://doi.org/10.4324/9780203154472

Vygotsky, L. S. (1987). Thinking and speech. In L. S. Vygotsky, R. W. Rieber (Series Eds.), & A. S. Carton (Vol. Ed.), *The collected works of L. S. Vygotsky. Vol. 1: Problems in general psychology* (N. Minick, Trans.). New York: Plenum.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The panas scales. *Journal of Personality and Social Psychology, 54*(6), 1063. https://doi.org/10.1037/0022-3514.54.6.1063

Wilkerson, M. H., & Polman, J. L. (2020). Situating data science: Exploring how relationships to data shape learning. *Journal of the Learning Sciences, 29*(1), 1–10. https://doi.org/10.1080/10508406.2019.1705664

Wong, R. M., & Adesope, O. O. (2020). Meta-Analysis of emotional designs in multimedia learning: A replication and extension study. *Educational Psychology Review, 33*(2), 357–385. https://doi.org/10.1007/s10648-020-09545-x