ARTICLE

Measurement of Science Museum Visitors' Emotional Experiences at Exhibits Designed to Encourage Productive Struggle

SARAH MAY, KATIE TODD, SAMANTHA G. DALEY, AND GABRIELLE RAPPOLT-SCHLICHTMANN

Correspondence
Sarah May,
Research & Evaluation Department,
Museum of Science, 1 Science Park,
Boston, MA 02114, USA.
Email: smay@mos.org

Abstract This study measured science museum visitors' emotional experiences at exhibits designed to encourage productive struggle (PS), an emotional experience where visitors productively persist through challenge. The study included 105 youth visitors ages 10–17 who interacted with one of three science exhibits designed to promote PS. Emotion was measured using a convergent mixed-methods design, including self-report ratings of the extent to which 13 emotions were experienced, coupled with narrative data generated through guided recall activities. Results indicated that most visitors experienced the constellation of emotions characteristic of PS, and that experiences of PS took on varied forms. These findings validated and enriched the emotional components expected to comprise PS, and suggested that the construct of PS can meaningfully characterize patterns of effortful engagement. Attending to visitors' emotional meaning-making, particularly through the productive resolution of challenge, can inform the purposeful design of exhibits that harness emotions to deepen visitor learning and engagement.

INTRODUCTION

Informal learning experiences are inherently emotional, and recent trends in museum design have attended to visitor emotions in increasingly nuanced ways (Norris & Tisdale, 2017; Rappolt-Schlichtmann et al. 2020). Emotion and learning are deeply intertwined (Immordino-Yang et al. 2018), and affective experiences can both foster engagement and support museum visitors to make meaning of their interactions with museum offerings (Falk & Dierking, 2000). However, prior research in informal science learning (ISL) has

Sarah May (smay@mos.org) is a Senior Research and Evaluation Associate at the Museum of Science, Boston, in Boston, Massachusetts, USA.
Katie Todd is the Director of Learning and Research at the Children's Museum of Pittsburgh, in Pittsburgh, Pennsylvania, USA.
Samantha G. Daley is an Assistant Professor in Counseling & Human Development, Educational Leadership, and Teaching & Curriculum at the Warner School of Education, University of Rochester, in New York, USA.
Gabrielle Rappolt-Schlichtmann is Executive Director and Chief Learning Scientist at EdTogether, Inc. in Boston, Massachusetts, and serves as an Adjunct Lecturer at the Harvard Graduate School of Education in Cambridge, Massachusetts, USA.

Curator: The Museum Journal 2021, 64.4 1–25 DOI: 10.1111/cura.12449

overemphasized the role of positive emotions (Staus & Falk, 2017), even though museum visitors value struggle and persist longer at exhibits when they engage in difficult tasks or new ideas (Allen, 2004; Packer & Ballantyne, 2002; Paris, 1997; Perry, 2012). Further, science museums themselves are increasingly required to address challenging socio-scientific topics (Pedretti & Navas Iannini, 2020b). Previous research conducted at the Museum of Science, Boston (MOS), a large US science museum, found that some visitors experienced negative emotions like confusion and frustration, and that (in certain cases) these emotions were associated with deeper engagement and overall feelings of satisfaction (Rappolt-Schlichtmann et al. 2017). This constellation of emotions can be described as *productive struggle* (PS), an emotional arc characterized by persistence through a challenging task, toward a satisfying end. To enhance the ability of designers to address such complex emotional engagement, new research and development models are needed that integrate current understandings of affective science within naturalistic ISL contexts.

To address this need, researchers at MOS conducted a multi-year design-based research (DBR) project. DBR involves the "close study of learning as it unfolds within a naturalistic context that contains theoretically-inspired innovations" (Barab, 2014, p. 151). As such, the project's objectives were to both develop and clarify understandings of PS in ISL contexts. Specifically, through this work, the team iterated on both the conceptual frameworks guiding PS design, and the development of PS science exhibits themselves. The objectives of the project were to articulate the emotional pillars comprising PS experiences, identify exhibit design strategies that elicit PS, use and test these design strategies in the development of science exhibits, and assess the learning and engagement outcomes associated with PS (Paneto et al. 2021).

Ultimately, three PS exhibits were developed, specifically designed with and for youth ages 10–17, and were each installed within the exhibit halls at MOS. The first two exhibits, *Sneak* and *Mystery Skulls*, were pre-existing exhibits identified by the team as needing refurbishment. Both exhibits were selected because they were hypothesized to foster experiences of PS to some extent, but it was expected that intentional re-design through a PS lens could benefit the visitor experience. Further, by refurbishing existing exhibits, the team was able to compare results of design choices between the original exhibit and revised prototypes. The third exhibit, *Air*, was developed from scratch to test design strategies emergent through the DBR process during development of the first two exhibits. These exhibits were selected based off their distinct visitor experience goals and to represent a range of science content areas commonly found in science and natural history museums, with an intention of developing design strategies applicable to a range of ISL contexts. For each exhibit, DBR took the form of iterative cycles of prototyping and testing, and then improving the exhibits and clarifying the guiding design strategies and overall PS design framework.

Sneak was the first exhibit the team refurbished to foster PS more intentionally. The final Sneak exhibit is a physical challenge in which visitors practice skills of observation and self-regulation as they learn natural history content about animal behavior. The main task is to move down a corridor slowly enough to sneak up on one of two virtual birds (an "easier" robin, or a "harder" wood thrush).

If the visitor goes too fast, the bird flies away and the activity ends; a display encourages the visitor to try again. When the visitor successfully self-regulates their speed as they approach the bird, the bushes on the screen part to show a scene of deer, and the visitor learns that deer listen to birds for alerts about approaching danger.

Mystery Skulls ("Skulls") was the second exhibit developed, and combines physical and digital elements to challenge visitors' animal classification skills. A screen directs visitors to select one of five skulls (physically mounted on turntables) and prompts visitors to guess what animal they think the skull belongs to, choosing between three options. After making their initial guess, visitors answer guided questions, with limited information, to gather evidence about skull features (e.g., whether eyes are on the front or sides) and how these features relate to behavior in the wild (e.g., predators typically have eyes on the front of their skull), that they ultimately use to confirm or disconfirm their initial prediction.

Air is a collection of activities that highlight surprising characteristics of air, and was developed from scratch to fully leverage the PS design framework from initial conception. The activities involve using fans to direct airflow in order to move ping-pong balls to various targets. Signage encourages, but does not require, visitors to start at "Air Basics," a set of simple demonstrations of unintuitive air properties. Beyond "Air Basics," a set of more complicated air activities prompt visitors to use phenomena such as vacuums and the Coandă effect to solve challenges.

To assess the results and impacts of the exhibits developed through this DBR process, a summative research study with 105 youth engaging with the exhibits explored three strands of inquiry: (1) whether visitors experienced the expected emotional arc of PS; (2) how exhibit design strategies supported PS; and (3) the extent of visitors' learning and engagement at these PS exhibits. The present article focuses on findings from the first strand, addressing two research questions:

RQ1. To what extent do visitors' descriptions of their emotional experiences with exhibits designed to support PS reflect the expected PS emotions?
RQ2. What temporal patterns of emotional experiences are evident at exhibits designed for PS?

Theoretical Basis for PS

Several psychological and developmental theories address the potential mechanisms behind experiences of struggle and productivity during learning. For example, Vygotsky (1980) described the zone of proximal development (ZPD), conceptualized as a learning space bounded by what one is currently able to do and what one can achieve with assistance from a more knowledgeable person. Nakamura and Csikszentmihalyi (2002) offered their theory of flow, which involves a "complete absorption in what one does" that occurs in intrinsically motivating activities that are appropriate levels of difficulty (p. 89). These theories present organizing frameworks in which we can position different types of learning experiences or emotional arcs. Like these frameworks, we conceptualize

PS as including emotional features, though we believe the specific emotional arc of PS is distinct. The concepts of cognitive dissonance (Festinger, 1962), desirable difficulty (Bjork & Bjork, 2011), hard fun (Papert, 2002), and productive failure (Kapur, 2008) each share common attributes with PS. Each has informed the team's conception of PS, though we have distinguished PS as distinct from these concepts by addressing its emotional components explicitly, embracing the role of negative emotions such as frustration, and conceiving of PS experiences from the perspective of learners in free-choice ISL contexts (Paneto et al. 2021).

Leveraging current research in the affective sciences, coupled with evidence gathered from visitor experiences through the iterative exhibit development process (described in Paneto et al. 2021), the team generated an operational definition for characterizing PS experiences. PS was defined as an emotional experience where learners: (1) encounter a challenging task and feel *disequilibrium*, (2) are supported to *persist* in the task and, (3) achieve a *productive* resolution. When they occur together, these three emotional phases (disequilibrium, persistence, and productivity) constitute a *full arc* of PS. Here we unpack the conjectured emotional characteristics of each of the three arc phases, integrating research from affective and learning sciences with knowledge built through our DBR work.

Disequilibrium

Disequilibrium includes emotions elicited when confronting impasses or obstacles during learning, and has been studied widely in formal learning contexts (D'Mello & Graesser, 2012; Lodge et al. 2018). Integrating such prior research with findings from our DBR studies on exhibit prototypes, we conceptualized that feelings of disequilibrium in ISL contexts could emerge during the course of learning when an experience feels novel or leverages uncertainty. As ISL contexts are inherently social, we also considered social unease as a source of disequilibrium (e.g., during competition or when dealing with interpersonal differences; Järvenoja & Järvelä, 2009; Pekrun & Linnenbrink-Garcia, 2012). Based on this theoretical grounding, we hypothesized that visitor descriptions of disequilibrium might incorporate several common emotion labels, including *confusion, frustration, surprise, challenge, disappointment*, or *nervousness*. Importantly, confusion and frustration can benefit deeper learning and engagement, particularly when learners have support to *persist* (D'Mello et al. 2014; Graesser et al. 2010).

Persistence

Under the right conditions, experiencing disequilibrium can enhance motivation to persist through effortful problem-solving (D'Mello et al. 2014). Persistence is conceptualized as both emotional (e.g., feeling motivation or determination) and behavioral (e.g., engaging in self-regulation or focused effort; Simon et al. 2015; Velayutham et al. 2011). We therefore hypothesized that evidence of persistence might include visitors' expressions of emotional involvement (based on emotion labels including *motivated* or *determined*), or behavioral engagement (including *focusing* on a task or *trying again*). We further conjectured that when provided with appropriate supports (such as feedback, choice, and support to self-regulate) learners would persist through disequilibrium toward a *productive* resolution (a return to equilibrium).

Productivity

Resolving disequilibrium in learning can lead to feelings of pride, delight, or that rare "Eureka!" moment characteristic of scientific discovery (D'Mello et al. 2009). While productivity could result from successful task completion, we hypothesized that for some visitors satisfaction might also result from the *perceived* effort put into the activity (Dweck, 1986). Therefore, rather than assessing productivity based on task-completion or related behavior, we sought to understand visitors' productivity in terms of its emotional tenor, including states like *satisfaction*, *happiness*, and *pride*. Ultimately, we hypothesized that visitors' persistence through disequilibrium towards productivity might foster deeper learning and engagement than experiencing only positive or negative states alone (Pekrun, 2014; Rappolt-Schlichtmann et al. 2017; Staus & Falk, 2017). This is important not only *in the moment* at ISL exhibits, but potentially for *future engagement* in STEM; the experience of PS may be a critical emotional intelligence for STEM professionals who grapple with new ideas that challenge their current understandings on a regular basis (Sinatra et al. 2014; Thagard, 2002). While characterizing the outcomes of PS experiences is outside the scope of the current paper, our first step was to validate whether this hypothetical set of emotions were indeed integrated in meaningful ways in visitors' experiences at exhibits intentionally designed to foster PS.

Measuring Emotion

Methods for measuring emotion vary by theoretical approach. For example, the *classical theory of emotion* posits that emotions are innate and universal, defined by distinct, essential, and objectively observable physiological signatures rather than subjectively interpreted feelings (Barrett, 2013; Ekman, 1993). In contrast, the *theory of constructed emotion* views it as an emergent phenomenon arising from the interaction between physiology, conceptual knowledge, contextual information, and social dynamics (Barrett, 2017).

The present study leverages the constructed view of emotions. Here, we characterize emotion as a continuous internal interpretation of the external world informed by personal history, background knowledge, and culture (Barrett et al. 2007). In taking this perspective, we prioritized subjective feelings over subconscious processes (i.e., respect for visitors' self-reported emotion over observed or physiological measures), and we emphasized the importance of integrating temporal and contextually-grounded anchors in visitors' emotional meaning-making. In doing so, we sought to present a methodological approach for emotion measurement that resonated with the constructivist frameworks commonly leveraged to measure learning in informal education contexts (Bell et al. 2009; Falk, 2016; Hein, 1998).

METHODS

To gather data describing study participants and their emotional engagement, we used two primary data collection methods: (1) post-surveys to gather quantitative ratings of emotions and

demographic characteristics, and (2) post-experience guided recall activities to gather qualitative retellings of visitors' emotional engagement.

Participants

Participants were public visitors selected through continuous random sampling in the Museum, and random sampling from a database of youth with a range of disabilities who previously expressed interest in participating in studies at the Museum (this was to ensure that at least 10% of our sample were visitors with disabilities, which is representative of the Museum's typical audience). Informed consent and assent procedures were approved by the Museum's Institutional Review Board (IRB). Participants included 105 youth ages 10–17 who engaged with exhibits called Sneak (n = 36), Mystery Skulls (n = 35), and Air (n = 34). Table 1 describes the sample: 51% were female, with the average age of 12 years. Most visitors identified as White (74%), followed by Asian or Asian American (15%), Hispanic or Latinx (8%), Black or African American (6%), Alaska Native or American Indian (1%), or another race (3%). Fourteen percent identified as having a disability, and eleven percent were not native English speakers.

Table 1.
Sample description, disaggregated by exhibit (N = 105)

	Sneak	Mystery skulls	Air	Total
Average age	12.6 (SD = 2.1)	12.4 (SD = 2.2)	12.2 (SD = 1.6)	12.4 (SD = 2.0)
Gender				
Female	18 (33%)	18 (33%)	18 (33%)	54 (51%)
Male	17 (35%)	15 (31%)	16 (33%)	48 (48%)
Prefers not to respond or missing	1 (33%)	2 (67%)	0 (0%)	3 (3%)
Race				
White	26 (33%)	27 (35%)	24 (31%)	78 (74%)
Black or African American	1 (17%)	3 (50%)	2 (33%)	6 (6%)
Hispanic or Latinx	2 (25%)	2 (25%)	4 (50%)	8 (8%)
Asian American	6 (38%)	4 (25%)	6 (38%)	16 (15%)
Alaska native or American Indian	0 (0%)	0 (0%)	1 (100%)	1 (1%)
Other race	0 (0%)	1 (33%)	2 (67%)	3 (3%)
Prefers not to respond or missing	3 (60%)	2 (40%)	0 (0%)	5 (5%)
Identifies as having a disability				
Yes	4 (27%)	8 (53%)	3 (20%)	15 (14%)
No	28 (33%)	26 (31%)	31 (36%)	85 (81%)
Prefers not to respond or missing	4 (80%)	1 (20%)	0 (0%)	5 (5%)
Identifies as native English speaker				
Native English speaker	31 (36%)	31 (36%)	23 (27%)	85 (81%)
Not native English speaker	2 (17%)	3 (25%)	7 (58%)	12 (11%)
Prefers not to respond or missing	3 (38%)	1 (13%)	4 (50%)	8 (8%)
Total	36 (34%)	35 (33%)	34 (32%)	105 (100%)

Note Visitors could select more than one race category, so the sum is greater than 100%.

Measures of Emotion

Survey

After engaging with the exhibit, participants completed a post-exhibit survey, which, for the purposes of this work, we have named the Productive Struggle Emotion Assessment Survey for Youth (PS-EASY). To measure emotion, the survey included the question, "Did you feel the following emotions while doing this activity?" for 13 emotions: *frustrated, challenged, surprised, disappointed, nervous, confused, focused, determined, motivated, persistent, proud, satisfied,* and *happy*. The PS-EASY instrument was adapted for this study based on the Positive and Negative Affect Schedule (PANAS; Watson et al. 1988), emotion ratings used in prior work (Rappolt-Schlichtmann et al. 2017), and emotion labels generated from visitor data throughout the exhibit development process. Participants rated whether they felt the 13 emotion labels along the Activation Lab's 4-point scale (with anchors of *YES!*, *yes*, *no*, *NO!*; Vincent-Ruz & Schunn, 2018). This scale allowed for consistency across questions used in the study. To ease cognitive load for participants, the 13 emotion labels were clustered into three groups (rather than randomized), categorized by our initial conceptualized PS categories (though they were not labeled as such). Further, follow-up questions probing the reasons for feeling each emotion were integrated in the survey to assess aspects of the exhibit design that might have elicited each emotion; however, these data are not reported as part of the current study as questions about the role of exhibit design features are outside the scope of the current paper. The PS-EASY instrument is available as part of this article's supplemental materials.

Guided recall

One of two methods of guided recall were then used to collect qualitative descriptions of visitors' (N = 104) subjectively felt emotions over time: *video stimulated recall* (n = 54) and *emotion storyboarding* (n = 50).

A *video stimulated recall* activity (n = 54) was adapted from recall approaches used in formal and informal learning research, as such approaches can be an unobtrusive way to gain insights about cognitive processes that cannot be observed (DeWitt & Osborne, 2010) and are adaptable for naturalistic contexts (Lyle, 2003). For this study, a think-aloud approach was implemented in which video footage anchored participants' narration. After visitors completed the emotion experience survey, researchers re-played video footage of visitors' engagement with the exhibits, prompting visitors to recall what they were thinking and feeling over time. Visitors' narration was audio-recorded and transcribed verbatim.

An *emotion storyboarding* (n = 50) activity was adapted from methods used in the design field to characterize emotion experiences over time (Chung & Gerber, 2010; Truong et al. 2006), and from participatory techniques that allow participants to define and describe the pivotal elements of their own experience, potentially reducing researcher bias in interpretation (Cross & Warwick-Booth, 2016). For this study, researchers developed a storyboarding protocol in which a step-by-step recollection of the visitor's experience was documented on ordered cards with short phrases that guided the narrative. While participants engaged with the exhibits, researchers documented sequential events during the visitors' experience (e.g., listing in order the activities they attempted at Air). After visitors completed the survey, researchers invited the participant to review and elaborate on these

ordered events. Visitors named any emotions they felt during each event by using a list of emotion words (both related and unrelated to PS) printed on stickers, or adding their own emotion descriptions. Researchers probed for detail about the context surrounding visitors' thoughts and feelings over time, and documented visitors' explanations verbatim, resulting in an annotated storyboard.

Because the video recall approach invited more granular description, we selected one exhibit to apply this approach across all cases for consistency (Skulls, $n = 35$), and then used this method with fewer Sneak ($n = 13$) and Air ($n = 6$) participants to include a subset of these more intensive sessions in conjunction with the storyboarding cases (Sneak, $n = 22$; Air, $n = 28$). Having both options available also helped us moderate the burden of data collection efforts, and provided options for visitors who did not consent to being filmed. Both methods generated narrative data about visitors' subjectively felt emotions in time series, including visitors' descriptions of why they felt different emotions. For each activity, transcriptions were uploaded to Dedoose, an online qualitative analysis software.

Analysis

We used a mixed methods convergent design in which PS-EASY responses and narrative retellings provided complementary perspectives on visitors' emotional experiences. Both survey and recall data helped identify whether visitors experienced the emotions hypothesized to constitute PS arc phases. Statistical analysis of the survey data explored relationships between ratings for each emotion to assess the extent to which our hypothesized categories (disequilibrium, persistence, and productivity) held together. The recall data allowed us to consider the temporal nature of visitor experiences, and to assess whether arc phases were integrated in *full PS arcs* as cohesive experiences (rather than simply the presence of unrelated instances of each individual arc phase). Findings converged to help us characterize the presence of arc phases, as well as the patterns and relationships between them.

Responses to the emotional experience survey were analyzed using descriptive statistics to assess the extent to which visitors reported feeling the discrete emotions constituting PS arc phases, and then a principal components analysis (PCA) was conducted to identify underlying dimensions and consolidate individual emotion responses.

The guided recall data were analyzed using qualitative methods for semi-structured interviews (Campbell et al. 2013; Forman & Damschroder, 2008). First, a codebook was developed *a priori*, with code categories aligning with the phases described in the PS definition (disequilibrium, persistence, and productivity), as well as three categories that related to other types of experiences we expected to find (experiences of *orienting* to the exhibit, such as figuring out what to do initially; *other emotions* unrelated to PS; and *negating evidence of PS,* such as evidence of visitors giving up when confronted with disequilibrium). This process was cyclical and iterative, and the team revised the codebook and coding process as insights emerged from new cases and across the exhibit contexts (Table 2).

Decisions about unitization and excerpting were based on immersion in the data and discussions about what would be most useful for analysis. An excerpt was ultimately defined as a length of

Table 2.
Final codebook used for guided recall narrative data

Code	Definition and criteria
Orientation	Feelings related to how the exhibit invited (or disinvited) engagement. A visitor's narration should include description of how the visitor felt when first deciding to engage with the exhibit, figuring out what to do, figuring out how to use the exhibit, or figuring out the goal of the activity; descriptions of what was compelling (or repelling) about using the exhibit
Disequilibrium	Feelings related to the visitor encountering a challenging task, idea, or phenomenon, and experiencing an emotional imbalance. Disequilibrium emotions are not necessarily negative, but do describe feelings of emotional imbalance (such as confusion, surprise, or frustration). The visitor's narration should include disequilibrium-resonant emotion language, or an explicit description of their subjective feeling of disequilibrium, and not simply a description of the level of challenge (i.e., simply labeling the "harder" bird at Sneak during narration does not imply a subjective emotional experience of challenge or struggle)
Persistence	Feelings or described behaviors associated with persistence through a task, or with an idea or phenomenon. The visitor's narration should include persistence-resonant emotion language (e.g., "determined"), a non-emotive description of behavioral persistence (e.g., "I tried again and again"), descriptions of growing more confident or capable over time, or descriptions of the ways in which the visitor thought about or puzzled through the task, idea, or phenomenon. Persistence events could include cases where the visitor anticipated persisting even after the research study ended
Productivity	Feelings related to achieving a positive resolution. Productivity emotions describe feelings of emotional rebalance or the subjective feelings associated with the resolution of some tension, effort, or feelings of disequilibrium (such as pride, satisfaction, or relief). The visitor's narration should include productivity-resonant emotion language, and not simply a description of the behavior of achieving a goal (i.e., simply stating that a visitor "finished the harder bird" does not imply a subjective emotional experience of productivity)
Other emotion	Emotional experiences that are not related to orientation or to one of the productive struggle arc phases. A visitor's narration might include any number of emotions unrelated to productive struggle, including emotions related to the research process (e.g., feeling "rushed" because of the time limit)
Negating evidence of PS	Clear evidence that the visitor had an emotional experience that discouraged one of the productive struggle arc phases in consequential ways. Narration might include descriptions of disequilibrium or persistence followed by boredom, frustration, disappointment, and/or giving up (rather than productivity or further persistence). Mere omission of descriptions of arc phases does not constitute evidence that an arc phase was not experienced or was discouraged
Ambiguous excerpts	Excerpts for which no arc or other emotion codes could be applied because the narrative was too ambiguous. This includes excerpts in which visitors describe what they thought or did, but did not include any clear emotion language to describe how they felt; this also includes excerpts in which emotions are labeled, but without enough context to confidently describe it as part of an arc phase or other emotion experience

dialogue that included enough detail to describe: (1) what a participant felt, and (2) why they felt that way or towards what event. Sometimes an excerpt only had one of the elements listed above (e.g., a participant described emotions, but it was impossible to know the antecedent or contextual referent to the emotions). We still created an excerpt, but in these cases we would apply a code of *ambiguous*. Sections of text that were irrelevant to the core objectives of the recall activity (e.g., visitors discussing the research equipment) were left uncoded.

As a core objective of this work was to both validate and clarify our understanding of the experience of PS, the team engaged in a negotiation and consensus building process for coding, in which reflexivity

and reason-giving were prioritized over achieving a pre-specified level of agreement independently (Forman & Damschroder, 2008). Transcripts were excerpted and coded in Dedoose by a primary coder who wrote short memos to explain their coding rationale, which were then reviewed by a secondary coder. At least two coders reviewed all transcripts multiple times, identified questions or areas of disagreement for discussion, and worked towards consensus on either a final code application or adjustments to the definitions. More difficult disagreements were brought to the full research team for input. This process of negotiated agreement was appropriate for this study, as it required coding that was sensitive to subtle meanings in the text, and included coders who had different levels of knowledge of the data (Campbell et al. 2013). Ultimately, coders came to agreement on all codes applied in the transcripts.

As a final step in the coding process, the team chunked contextually connected series of excerpts. For example, a series of excerpts in which a visitor described their engagement with the robin at the Sneak exhibit were chunked, followed by a separate chunk for a subsequent series of excerpts in which they described attempting the wood thrush. This chunking helped the team analyze whether and how different emotions were related to visitors' context (in this case, which bird they were trying) and behavior (such as trying a new bird, or repeating one of the birds). This final chunking of the data allowed the team to uncover patterns in emotional experiences over time (such as whether the arc phases described were meaningfully interrelated). When identified, such patterns were discussed and interpreted by the team, and each participants' transcripts were coded for the presence of these emergent patterns.

RESULTS

Emotion Descriptions

To address RQ1, we examined patterns of emotional experiences in the quantitative and qualitative data as they related to our hypothesized PS phases. Descriptive analysis of the survey data sought to uncover the extent to which visitors experienced the listed emotions. PCA analysis assessed whether emotions reported by visitors hung together as distinct component groupings across cases, and whether these component groups aligned conceptually with hypothesized PS phases. Qualitative data were examined for contextual factors and other aspects of the narrative retellings that might validate whether PS phases were experienced as hypothesized. While the presence of different emotion labels in the retellings were examined during analysis, frequencies of specific emotion labels were not calculated since the number of emotion words used was not an accurate indicator of the extent to which visitors might have experienced these feelings (e.g., sometimes visitors repeated an emotion word several times when describing just a single event).

PS Emotion Components

We expected that visitors to PS exhibits would experience the three PS phases: disequilibrium, persistence, and productivity. Survey items included ratings on 13 emotions we hypothesized to relate

conceptually to each of these concepts (Table 3). From this survey data we found that most visitors reported feeling at least one of the emotions we expected to characterize disequilibrium (96% *yes* or *YES!*), persistence (96% *yes* or *YES!*), and productivity (99% *yes* or *YES!*).

A PCA further confirmed that the hypothesized emotions constituting these categories were indeed meaningfully related. Interestingly, three subtypes of the disequilibrium category emerged for a total of five components (Table 4): (1) Persistence; (2) Productivity; (3) Challenge Disequilibrium; (4) Frustrated Disequilibrium; and (5) Nervous Disequilibrium. Table 5 presents the loadings using oblimin rotation. At least 60% of the variance for each survey item was explained (average of 73%) by these five categories, which is an acceptable range for explaining such variance in the social sciences (Hair et al. 2014). This means that we could meaningfully condense the observed ratings for each of the 13 individual emotions into these five overarching categories, each of which we found conceptually useful for characterizing PS. An exploratory factor analysis confirmed the robustness of the PCA results, finding similar results and further supporting both the *a priori* categorization of emotions and the finer-grained nuance within disequilibrium.

Disequilibrium

A content analysis of the narrative recall data supported findings from the survey data, revealing that descriptions of emotional experiences aligning with each PS phase were present in most participants' retellings (Table 6).

Descriptions of disequilibrium emerged in 74% of Sneak cases, 94% of Skulls cases, and 97% of Air cases. We expected confusion, surprise, and frustration to be characteristic of

Table 3.
Summary of responses to the PS-EASY item, "Did you feel the following emotions while doing this activity?"

Emotion	*NO!* (%)	*no* (%)	*yes* (%)	*YES!* (%)	Total *yes* or *YES!*
Disequilibrium					
Frustrated	31 (30.4)	40 (39.2)	27 (26.5)	4 (3.9)	98 (96.1)
Challenged	1 (1.0)	13 (12.7)	62 (60.8)	26 (25.5)	
Surprised	2 (2.0)	25 (24.5)	47 (46.1)	28 (27.5)	
Disappointed	40 (39.2)	49 (48.0)	12 (11.8)	1 (1.0)	
Nervous	32 (31.4)	51 (50.0)	17 (16.7)	2 (2.0)	
Confused	13 (12.7)	38 (37.3)	38 (37.3)	13 (12.7)	
Persistence					
Focused	1 (1.0)	7 (6.9)	54 (52.9)	40 (39.2)	98 (96.1)
Determined	2 (2.0)	11 (10.8)	57 (55.9)	32 (31.4)	
Motivated	2 (2.0)	16 (15.7)	53 (52.0)	31 (30.4)	
Persistent	3 (2.9)	10 (9.8)	54 (52.9)	35 (34.3)	
Productivity					
Proud	2 (2.0)	8 (7.8)	60 (58.8)	32 (31.4)	101 (99.0)
Satisfied	0 (0.0)	4 (3.9)	56 (54.9)	42 (41.2)	
Happy	2 (2.0)	5 (4.9)	57 (55.9)	37 (36.3)	

Note $N = 102$. Three participants did not complete the survey. The categorizing labels for "Disequilibrium," "Persistence," and "Productivity" were not incorporated in the survey.

Table 4.
PS-EASY items comprising PCA components and conceptual labels

Component	Survey items	Conceptual label
1	Determined, motivated, persistent, focused and NOT confused	Persistence
2	Proud, satisfied, happy	Resolution
3	Challenged, surprised, confused	Challenge Disequilibrium
4	Frustrated, disappointed	Frustrated Disequilibrium
5	Nervous, NOT satisfied	Nervous Disequilibrium

disequilibrium, and visitors indeed used these terms when discussing their experiences. Other visitors tended to use less obvious descriptive language to imply the emotional experience of "not knowing," such as one visitor describing his reaction when observing a skull's features, "I was like, if [the teeth are] sharp, *I have no clue*. If they're more flat, *I have no clue*." Further, we found that emotion labels such as "unsure," "uncertain," and others beyond our original list were helpful for identifying descriptions of disequilibrium. For example, when discussing her attempt at the harder bird at Sneak, one visitor described how her emotional state shifted compared to trying the easier bird, stating, "I felt *skeptical* and *hesitant* because it was different and it didn't show the sneak meter like for the robin. I was *unsure* how to do it." Another visitor described her experience at Air using both traditional and more contemporary emotion descriptions, saying, "I felt

Table 5.
PS-EASY items, their means and standard deviations, and loadings for principal components analysis with oblimin rotation

		Components				
Item	Mean (*SD*)	Component 1: Persistence	Component 2: Resolution	Component 3: Challenge disequilibrium	Component 4: Frustration disequilibrium	Component 5: Nervous disequilibrium
Frustrated	2.04 (0.86)	0.03	0.02	0.09	**0.63**	0.12
Challenged	3.11 (0.64)	0.04	−0.18	**0.63**	0.05	−0.11
Surprised	2.99 (0.78)	0.01	0.08	**0.55**	−0.13	0.07
Disappointed	1.75 (0.70)	−0.03	−0.04	−0.07	**0.66**	−0.04
Nervous	1.89 (0.74)	0.02	0.04	−0.02	0.04	**0.86**
Confused	2.50 (0.88)	−**0.48**	0.19	**0.41**	0.20	0.04
Focused	3.30 (0.64)	**0.25**	0.21	0.15	−0.10	−0.02
Determined	3.17 (0.69)	**0.44**	0.00	0.22	0.00	0.01
Motivated	3.11 (0.73)	**0.48**	0.01	0.13	−0.06	0.22
Persistent	3.19 (0.73)	**0.51**	0.05	−0.05	0.23	−0.10
Proud	3.20 (0.66)	0.02	**0.61**	−0.13	0.08	0.10
Satisfied	3.37 (0.56)	0.14	**0.37**	−0.00	0.13	−**0.38**
Happy	3.28 (0.65)	−0.08	**0.61**	0.05	−0.14	0.02
Percentage of variance		0.19	0.17	0.15	0.13	0.09
Eigenvalue		4.48	2.25	1.15	0.88	0.73

Note Factor loadings >0.25 are in boldface. This highlights items that are highly correlated with each principal component. Range is 1 (*NO!*), 2 (*no*), 3 (*yes*), and 4 (*YES!*).

Table 6.
Frequency of visitors reporting PS phases and other emotion experiences, total and disaggregated by exhibit (N = 104)

	Orientation (%)	Disequilibrium (%)	Persistence (%)	Productivity (%)	Other emotion (%)	Ambiguous excerpt(s) (%)	Negating evidence of PS phase (%)
Sneak (n = 35)	25 (71.4)	26 (74.3)	32 (91.4)	32 (91.4)	16 (45.7)	10 (28.6)	2 (5.7)
Skulls (n = 35)	25 (71.4)	33 (94.3)	32 (91.4)	31 (88.6)	28 (80.0)	29 (82.9)	2 (5.7)
Air (n = 34)	24 (70.6)	33 (97.1)	34 (100.0)	30 (88.2)	27 (79.4)	26 (76.5)	6 (17.6)
Total (N = 104)	74 (71.2)	92 (88.5)	98 (94.2)	93 (89.4)	71 (68.3)	65 (62.5)	10 (9.6)

frustrated because I did it at the beginning, but couldn't do it again. When the ball bounced out, I felt *shook*." Descriptions of disequilibrium varied between visitors and across the exhibit contexts, but ultimately the concept of disequilibrium hung together across diverse descriptions through retellings that invoked a sense of emotional imbalance. Whether perceived as positive (surprise) or negative (frustration), these imbalances were characterized by the sense that the visitor might need to follow up or push ahead in order to resolve a lingering tension or overcome an obstacle in their exhibit progress.

Persistence

Descriptions of persistence emerged in 91% of Sneak cases, 91% of Skulls cases, and 100% of Air cases. Descriptions of persistence were more variable than descriptions of disequilibrium and productivity, due to our definition of persistence as including both affective and behavioral elements. Some visitors indeed described persistence using expected emotion labels, such as "motivated," "determined," and "persistent." An emergent characteristic of persistence arose as we found visitors who described how the exhibit supported them to build confidence over time. For example, one visitor initially described feeling "nervous" and "unsure" making guesses at Skulls, but then she explained how the exhibit supported her persistence and accompanying increase in confidence, saying, "I got to break it down. Breaking down the activity really made me feel more *confident*." Another emergent theme came from visitors narrating their internal problem-solving process. For example, another visitor to Skulls described his uncertainty before having to make an initial guess, explaining, "That one I had no idea. At first I thought it was a rattlesnake." Instead of simply making his first guess based on this hunch, the visitor went on to explain his more in-depth reflection and decision-making process, indicative of his direct engagement through feelings of disequilibrium. He recalled his observations, "Right here [the head] just kind of expands and that's what the rattlesnake head looked like. But the bumps at the top made me think it was the Gila monster." Later he reiterated, "That was probably the most difficult out of all of them." Whether through emotion descriptions, or retellings involving behavioral persistence or decision-making through disequilibrium, excerpts in this category shared a

sense of forward momentum and initiative, either towards a final goal or simply past an immediate obstacle.

Productivity

Finally, descriptions of productivity emerged in 91% of Sneak cases, 89% of Skulls cases, and 88% of Air cases. Descriptions of productivity often included expected emotion labels, like "satisfaction" or "pride," that strongly implied the experience of some resolution. A sense of "relief" or "accomplishment" were also common. One visitor stated, "I think I would use the word *relieved* actually! . . . Because I finally did it." In other cases, contextual information was necessary to couple more ambiguous emotion labels, like feeling "happy" or "good," with the experience of productive resolution. After attempting and accomplishing several tasks at Air, one visitor described, "I felt *happy* because I did it." In some cases, visitors used non-emotive language that still communicated something about their affective experience resolving a challenge. Visitors at Skulls and Air described feeling "smart" after achieving their goals. One visitor at Sneak characterized her emotional state with the exclamation, "*Oh! I finally did it!*" Finally, beyond experiences of completing a task or achieving an exhibit-defined goal, we also found evidence of productivity from some visitors' personally defined sense of accomplishment. For example, one visitor at Air felt "*proud* because I got close, and at least I tried." Another felt productivity for social reasons, stating, "I was *satisfied* because I beat my brother." Overall, the common thread in visitors' retellings of productivity was the sense that emotional tension was resolved, whether through a heightened emotional resolution such as pride or accomplishment, or through more subdued emotions like relief or feeling pleased.

Negating evidence of PS

We found some evidence of moments during visitors' experiences where movement through a PS phase was obstructed. While few in number, characterizing these experiences was useful to assess whether exhibits designed for PS might – for some visitors – actually discourage the emotions we sought to foster. In all, 17 excerpts across 10 visitors' retellings (9.6% of participants) included evidence that PS phases were discouraged. A majority of these were at the Air exhibit (6 of 10), with Skulls (2 of 10) and Sneak (2 of 10) sharing the remainder. In all excerpts, visitor descriptions were related to perceived levels of disequilibrium. In 13 of the 17 excerpts, visitors described how enduring or intense disequilibrium led them to quit rather than persist towards a productive resolution. For example, a visitor at Air explained her reason for moving to the next activity, saying, "I had tried so much and *couldn't figure it out*, so I thought I should move on and try out the other stuff." In four cases, visitors described boredom from NOT feeling enough challenge. For example, one visitor to Skulls suggested, "It was a little *boring*. . . because it seemed to be stuff that I already knew." This evidence suggests that as many as 10% of visitors might experience undesirable levels of disequilibrium – either too much or too little – at some point during their experience with PS exhibits.

Non-PS emotions

The narrative data suggested that other emotions were also common. During the recall interviews, many participants (80.0% at Air; 79.4% at Skulls; 45.7% at Sneak) described emotional

experiences that did not fall within the scope of PS (Table 6). The emotions tended to vary by exhibit (Table 7). At Air, respondents most often described feeling positive, active states like having fun (15 of 34), feeling happy or satisfied (13 of 34), being interested (11 of 34) or excited (9 of 34). Although happiness and satisfaction can be indicators of productivity, the 13 visitors described these emotions as distinct from productivity. For instance, one participant talked about feeling happy to be around other people. At Skulls, the most frequent emotional experiences outside of PS was feeling confident (18 of 35 participants) at the *outset* of an activity (unlike the confidence coded as persistence described previously, in which visitors described *gaining* confidence through persistence). Visitors to Skulls also felt that the exhibit was easy (11 of 35) and reported being calm (8 of 35). Participants tended to report the fewest emotions at Sneak. The only emotion outside the PS arc that more than five Sneak visitors shared was feeling happy (6 of 35). As mentioned above, happiness can be an indication of productivity, but visitor descriptions suggested they were not in these cases. For instance, one described, "I watched my sister do the robin. I felt *happy* and *satisfied* because she did it, too."

Emotional Arc Patterns

We addressed RQ2 by examining patterns of visitors' emotional experience over time based on their guided recall narratives. After chunking excerpts by visitors' context and behavior, we assessed whether the three arc phases were interrelated in meaningful ways. We also characterized refuting evidence (such that PS was explicitly *not* evident) in emotional experiences, including evidence of partial PS arcs and any evidence that PS arcs were problematically discouraged.

Complete PS arcs

Most visitors experienced at least one full arc of PS (77.9% of participants). Descriptions of full arcs tended to fall into three types: mini-arcs, extended arcs, and emergent arcs (Table 8). These could occur any number of times, and there could be two or more arc types in a single narrative.

Mini-arcs were contextually-bounded experiences in which visitors engaged with disequilibrium, persisted, and came to a productive resolution in sequence and without added variation of other non-PS emotions. For example, one visitor described a bounded experience at an Air activity. She began describing an experience of disequilibrium, "I tried Vacuum Maze, and I felt frustrated. . . because it

Table 7.
Frequency of visitors reporting other emotion sub-categories, disaggregated by exhibit (N = 104)

	Ease (%)	Confidence (%)	Fun (%)	Happiness (%)	Interest (%)	Calm (%)	Excitement (%)
Sneak (n = 35)	3 (8.6)	0 (0.0)	2 (5.7)	6 (17.1)	0 (0.0)	2 (5.7)	0 (0.0)
Skulls (n = 35)	11 (31.4)	18 (51.4)	3 (8.6)	1 (2.9)	4 (11.4)	8 (22.9)	2 (5.7)
Air (n = 34)	8 (23.5)	3 (8.8)	15 (44.1)	13 (38.2)	11 (32.4)	2 (5.9)	9 (26.5)

Table 8.
Frequency of visitors who narrated full PS arcs and sub-types, total and disaggregated by exhibit (N = 104)

	Sub-types			
	Full PS arc(s) (%)	**Mini arc(s) (%)**	**Extended arc(s) (%)**	**Emergent arc(s) (%)**
Sneak (n = 35)	22 (62.9)	12 (34.3)	9 (25.7)	13 (37.1)
Skulls (n = 35)	29 (82.9)	26 (74.3)	1 (2.9)	16 (45.7)
Air (n = 34)	30 (88.2)	21 (61.8)	20 (58.8)	11 (32.4)
Total (N = 104)	81 (77.9)	59 (56.7)	30 (28.8)	40 (38.5)

was harder than Direct. The ball didn't want to move," followed by persistence through the disequilibrium, "Once I was so close, but it pushed the ball away. Made me mad. I had to figure out other ways to do it," and ending with productivity, "I was proud because I got close, and at least I tried." Over half the participants described mini-arcs.

Extended arcs of PS occurred when visitors engaged with disequilibrium early within a particular context, but then experienced some emotional variability or spent time trying something else at the exhibit before eventually resolving the original disequilibrium. For example, one visitor described intentionally spending time away from a particularly difficult Air activity, explaining, "I had reached a stalemate. I decided to do other activities to refresh my psyche, see if I could get a lightbulb to pop up." He later returned to try again, reaching a productive resolution, "Compared to the last time, I figured out how it worked and felt happy." Over a quarter of participants described extended arcs of PS, emphasizing that some visitors might seek or require different supports to navigate (or even avoid) feelings of disequilibrium prior to re-engaging and resolving challenges productively.

In *emergent arcs* of PS, rather than describing productivity in terms of resolving a specific instance of disequilibrium, visitors narrated an overarching sense of productivity after persisting through disequilibrium and engaging with the exhibit as a whole. For example, when the interviewer asked one visitor, "How did you feel when you finished the exhibit?" she responded, "Very, like, satisfying, 'cause now I have, like, learned some new stuff." Another visitor at Sneak stated that she was "proud to accomplish both birds." While these visitors used the *culmination* of their experiences to describe productivity, some emergent arcs of PS arose from visitors' retellings that indicated they were continuously *comparing* feelings of productivity to assess which aspects of their exhibit experience were most meaningfully productive. For example, one visitor at Air compared two activities, explaining:

> *I tried Direct and I felt not quite proud, just less than that. . . It was cool to beat it, but then I just moved on. Like, I wouldn't show it off to my dad or brother. . . . Next I went to Hover Pass again, and I felt proud and overjoyed. When I beat it, I showed my dad and brother.*

With over a third of visitors describing experiences of emergent arcs of PS, these results highlighted the nature of some PS experiences that do not simply revolve around the completion of tasks or achievement of goals. In these cases, visitors can walk away from a PS exhibit with a more holistic sense of productivity.

Table 9.
Frequency of visitors who narrated partial arcs and sub-types, total and disaggregated by exhibit (N = 104)

	Sub-types		
	Partial arcs (%)	**Disintegrated phases (%)**	**PS discouraged (%)**
Sneak (n = 35)	13 (37.1)	4 (11.4)	1 (2.9)
Skulls (n = 35)	6 (17.1)	4 (11.4)	0 (0.0)
Air (n = 34)	4 (11.8)	8 (23.5)	4 (11.8)
Total (N = 104)	23 (22.1)	16 (15.4)	5 (4.8)

Incomplete PS arcs

Not every visitor described a full PS arc (Table 9). In some cases (22.1% of participants), visitors omitted one or more of the three arc phases from their full narratives (*partial arcs*). We also documented cases (15.4% of participants) where chunks within visitors' narrative included all three phases, but these were either narrated out of order or the emotions occurred in response to unrelated events, suggesting the mere presence of all three phases was not sufficient evidence of a fully interrelated PS arc (*disintegrated arcs*). It is important to note that visitors who described *disintegrated arcs* often also described full PS arcs during other parts of their experience (this was true for 12 of these 16 cases; the remaining 4 cases were included in the *partial arc* case counts).

Discouraged PS arcs

Several visitors who experienced obstructed arc phases (described above) did ultimately experience full PS arcs at some point (n = 4 of 10). However, in four Air cases and one Sneak case, we found evidence that a full experience of PS was *discouraged* in problematic ways. These cases included instances of negating evidence of PS arc phases that ultimately resulted in a visitor exiting an exhibit feeling frustrated or incapable of a productive resolution. For example, one visitor stated he left an Air activity because, "I tried it like a lot of times and I just knew I wouldn't get better at it." At Sneak, one visitor described experiencing increasingly negative emotions, from upset to defeated, before she explained, "I couldn't figure out how to go even slower. I gave up." While few in number, these cases highlight the potential for PS exhibits to elicit unintended discouragement.

DISCUSSION

Informal science learning environments have the potential to invite visitors into complex emotional journeys where disruption, determination, and delight all come into play. Extant theoretical frameworks have supported educational experience designers to attend to struggle in meaningful ways, but the current work sought to address *productive* struggle by explicitly addressing the emotional aspects of PS during visitors' experiences of it. Through a convergent mixed-methods design that prioritized visitors' self-reported emotional experiences, we sought to generate a contextually-grounded understanding of PS as a construct of emotional engagement that could inform exhibit design practice. We indeed found that visitors' descriptions of their emotional experiences reflected

the expected PS emotions, with variability across PS experiences. Used together, the PS-EASY instrument coupled with qualitative guided recall measures successfully characterized visitors' emotions in ways that helped assess the extent to which visitors experienced PS, and helped enrich an understanding of the varied ways in which PS emerged for visitors. These findings illustrated that it is indeed possible for design to attend meaningfully to diverse learners' emotional experiences, with implications for ISL practitioners seeking to support learners intentionally through challenge and struggle at science exhibits.

Revisiting PS Emotions

For each distinct phase of PS, in both the survey and narrative data, most visitors reported emotional experiences that aligned with – and expanded – our team's original conceptions. Disequilibrium emerged for nearly all visitors, not only in terms of surprise, confusion, or frustration, but also as disruptions in visitors' experiences that they characterized in terms both contextually and culturally relevant to them. Visitors described feeling unsure, uncertain, skeptical, and even "shook" or having "no clue" when discussing their shifting feelings in the face of challenge, novelty, or an impasse at science exhibits. Interestingly, the qualitative nuance in visitors' descriptions of disequilibrium was reinforced in the quantitative analysis, which suggested sub-types of disequilibrium might be present – namely challenge, frustration, and nervousness. Future research might explore how such subtypes relate to exhibit design features, experiences of PS, or other outcomes. For example, we wonder whether the frustration sub-type might be more strongly associated with discouraged arcs of PS (particularly those that lead to premature disengagement, a critical issue in choice-based learning contexts), or if the nervous sub-type might be related to reduced STEM learning outcomes (Blanco et al. 2010; Earl et al. 2017). Overall, this finding suggests that ISL designers can indeed work to elicit stereotypically negative emotions from visitors, and do so in ways that also encourage persistence through and productive resolutions of such moments.

Visitor narratives also expanded our understanding of the emotional and behavioral evidence characterizing persistence. While visitors' emotional descriptions included specific labels like determined and motivated, a notable addition emerged when visitors described gaining confidence over time, which may have implications for the role of PS on self-efficacy in science learning (Rittmayer & Beier, 2009; Vermeer et al. 2000). Further, when describing why they kept trying exhibits, or how they addressed challenges or confusion, visitors' descriptions of their thinking and decision-making illustrated the diverse ways they engaged with disequilibrium head-on, even if it was not observably evident or associated with a specific emotional state. This finding resonates with prior work illustrating the complex relationship between motivation, behavior, and decision-making that feed into goal-directed processes like persistence (Meier & Albrecht, 2003). These findings suggest the importance of exhibit design that includes both emotional and cognitive support for visitors to persist through challenge.

Productivity described by visitors also helped us broaden our understanding of this phase. Beyond the active and positive emotions we initially generated and found agreement with in survey ratings (happy, satisfied, proud), we found evidence of a different type of release, namely relief, that emerged for some visitors. This finding aligns with literature that distinguishes activating achievement emotions (like enjoyment or pride) from deactivating achievement emotions (like relief or contentment) in the context of learning (Pekrun, 2011; Pekrun et al. 2002). While deactivating emotions can reduce motivation to continue making an effort, they might also reinforce motivation to return to learning material later (Pekrun, 2014). Notably, this tension aligns with our observation that some visitors indeed disengaged and then re-engaged with sources of disequilibrium during their time at exhibits. While visitor choice is inherent in ISL contexts, one implication of this particular finding is the importance of acknowledging that exiting an experience does not always mean a visitor has completely disengaged. Exhibit design can more intentionally incorporate opportunities for fluid disengagement and re-engagement to invite such use. Further, future research might explore the differences, if any, between activating and deactivating productivity emotions and their relationship with PS and associated outcomes.

Finally, we noted that non-PS emotions played a role in visitors' experiences. The presence of these emotional experiences warrant at least two interpretations. First, visitors bring with them diverse emotional inclinations, histories, and expressive habits. As such, this emotional variability could simply be an expected, yet *incidental*, outcome of studying visitors' emotional experiences using open-ended guided recall. Second, this finding suggests the potential for designing exhibits (and, further, multiple activities within exhibitions) such that non-PS experiences work *intentionally* to provide visitors with opportunities for emotional alleviation from more intensive PS experiences. Future studies could explore the impact of intentionally encouraging time away to process, reflect, or otherwise cope with aspects of PS.

Revisiting PS Arcs

Patterns of PS arcs also tended to align with our original conception, in that we found narrative evidence that the emotional experience of productivity was often tied to visitors' efforts to persist through disequilibrium. Here, too, we expanded our original conception of what a full PS arc entailed, as several PS types emerged. We had not conceptualized the three pillars of PS to necessarily function linearly, as we understood that even *anticipated* satisfaction can lead to self-regulated persistence through struggle (Paris & Newman, 1990), and learners can engage in *cycles* of disequilibrium and persistence prior to successful resolution of a challenge (D'Mello et al. 2010). However, unpacking the differences between mini-arcs, extended arcs, and emergent arcs enabled us to articulate the temporal and sometimes cyclical dynamics of PS as it naturally occurred. These findings have implications for exhibit designers seeking to foster PS experiences: there is no one-size-fits-all approach to grooming emotional paths. Instead, design for PS should embrace the goal of supporting the most visitors possible to resolve disequilibrium productively, acknowledging the diverse routes or pace they

might take to get there. Such an approach to design is uniquely appropriate for ISL settings, which often prioritize choice and flexibility in visitor engagement.

Evidence that some visitors experienced *discouraged* PS phases or arcs, particularly in relation to experiences of disequilibrium, suggests that we might need to examine more closely the balance of disruptions and supports provided. Interestingly, the greatest portion of discouraged PS moments emerged at the Air exhibit, the final exhibit developed through this work and the only exhibit developed from scratch. One possible interpretation might be that as the team members became more comfortable engaging visitors with moments of disequilibrium, we also became less sensitized to eliciting strong negative emotions among these learners, and perhaps pushed too far in this final exhibit. Future research should address the individual or contextual factors that might be at play when PS is problematically discouraged. While we found no such difference along demographic variables from our sample, we have considered the potential need for a focused exploration of experiences of PS within individuals who are more likely to struggle with the management of intense emotions. For example, individuals with Attention Deficit Hyperactivity Disorder (ADHD), anxiety disorders, or Autism Spectrum Disorder (ASD) might navigate PS differently (Rappolt-Schlichtmann et al. 2020). Future work should consider emotionally accessible design for PS, such that the burden of inclusion is placed on the designed environment and not on visitors themselves.

Ultimately, evidence from the present study confirmed that most visitors to PS exhibits indeed experienced cohesive arcs of disequilibrium, persistence, and productivity. With an enriched perspective on the components within and patterns among PS arcs, we now summarize an elaborated definition of PS:

Productive struggle is an emotional experience where learners: (1) encounter a disruptive task, phenomena, or idea and shift into a state of *disequilibrium*, (2) are supported to *persist* through disequilibrium using emotional or behavioral resources (e.g., motivation, self-efficacy, problem-solving, trying again, etc.), and (3) achieve an emotionally *productive* resolution tied to the source of disequilibrium or a more holistic sense of effortful achievement. These three emotional pillars (disequilibrium, persistence, and productivity) can constitute a *full arc* of PS when visitors have flexible opportunities to persist through and resolve disequilibrium.

Limitations

The present research is not without limitations. We first acknowledge that all research methodologies – and even the construct of emotion itself – are culturally situated. We consider the methods and data interpretations put forward here as one perspective on this work; we welcome considerations emphasizing new or even contradicting interpretations in our findings. Methodologically, this study heavily relied on qualitative data from diverse visitors in a naturalistic museum context, leading to variability in the quality and quantity of narrative data collected. As such, we acknowledge that the presence or absence of evidence of emotional experiences might not represent the full range of

experiences visitors actually felt. Variability in communication styles and in researcher probing was assumed during interpretation of the findings, and should be acknowledged in further use of these findings. Further, we acknowledge that the PS-EASY survey design, while intended to streamline ease of use by visitors, potentially pre-confirmed our emotional phases of PS by clustering certain items together. This weakens our ability to interpret the PCA findings in this case, although convergent findings from our qualitative analyses do strengthen and validate our overall results. The PS-EASY tool itself, as an efficient instrument for assessing the extent to which visitors felt a targeted set of PS emotions, might be effectively adapted, piloted, and further validated in other learning and research contexts. In conjunction with guided recall measures, we believe this mixed-methods approach to emotion measurement can be used in future research to robustly characterize museum visitors' emotion experiences.

Finally, the present study was conducted with a relatively small number of visitors at a single museum, meaning any direct extrapolation to other contexts should be considered conservatively. However, the general framework for understanding and designing for PS was meant to be leveraged in other contexts, and as such we anticipate re-contextualized adaptations of this work should still be fruitful. For example, struggle does not only emerge in the context of engaging with challenging concepts or surprising phenomena, as the exhibits in the current study did. Socio-scientific issues, content addressing ethical dilemmas in science, or other socially controversial topics might introduce their own host of emotions associated with feelings of disequilibrium, thus requiring focused attention on the potential emotional arcs visitors might experience and intentional design strategies to support productive engagement with these feelings. Indeed, as Pedretti and Navas Iannini (2020a) suggest, productive struggle and productive discomfort can arise when visitors confront and engage with difficult topics, requiring "a need to balance feeling hopeless, fearful, or overwhelmed with feelings of hope and possibility" (p. 234). If the current study is any example, research leveraging the PS framework can help uncover and disentangle the interaction between design strategies and their associated emotional pathways in visitor experiences with such challenging topics. While the emotion labels and context might differ, understanding the potential variation in visitors' emotional arcs could inform the design of such controversial exhibits. For example, our own work suggested that some individuals benefit from time away from difficult experiences before re-engaging, while others drew on resources allowing them to persist more directly through their disequilibrium; both types of learners could eventually resolve their initial disequilibrium. How might controversial exhibits be designed to support both types of learners? By understanding the diverse emotional pathways visitors might follow, developers can more readily attend to design features that address the emotional needs of learners in personally relevant ways.

CONCLUSION

Attending to the complex dynamics of visitors' emotions in exhibit design can create rich opportunities for effortful informal learning experiences. Findings from the current study provided evidence that exhibits designed to foster PS can successfully support visitors through this complex

emotional experience. Overall, this work contributes to research and development in informal learning environments by first unpacking the nuance within experiences of a single emotional construct (PS), and then illustrating that understanding such nuance is integral to the design of exhibits that support diverse visitors through an intended emotional journey. Emotions are dynamic, personally and contextually grounded, and difficult to predict, but museum practice cannot make serious advancements without direct attention to visitors' complex emotions. At a time when science itself is increasingly questioned, science museums are beginning to position themselves as "important players in a number of external scientific, social, cultural, and political contexts" (Pedretti & Navas Iannini, 2020b, p. 704) and have begun to address the importance of working through "difficult ideas and conflicting beliefs in meaning-making and sense-making experiences" (p. 709). As such, attending intentionally to visitors' emotional meaning-making, particularly through the productive resolution of challenge, can inform the purposeful design of exhibits that do not just elicit emotions, but harness them to deepen visitor learning and engagement. END

ACKNOWLEDGMENTS

This study is based upon work supported by the National Science Foundation under Grant No.'s #1222613 and #1612577. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors thank the project team and our study participants for contributing to our evolving understanding of productive struggle, Becki Kipling for her support on this article, and Deborah Perry, J. Shipley Newlin, Josh Gutwill, Lauri Halderman, and Sidney D'Mello for providing feedback and guidance throughout the project. We would also like to thank our anonymous peer reviewers for providing comments and feedback that further strengthened this work.

REFERENCES

Allen, S. (2004). Designs for learning: Studying science museum exhibits that do more than entertain. *Science Education*, 88(S1), S17–S33. https://doi.org/10.1002/sce.20016

Barab, S. (2014). Design-based research: A methodological toolkit for engineering change. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 151–170). Cambridge University Press.

Barrett, L. F. (2013). Psychological construction: The Darwinian approach to the science of emotion. *Emotion Review*, 5(4), 379–389.

——— (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, 12(1), 1–23.

———, Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. *Annual Review of Psychology*, 58(1), 373–403. https://doi.org/10.1146/annurev.psych.58.110405.085709

Bell, P., Lewenstein, B., Shouse, A. W., & Feder, M. A. (Eds.) (2009). *Learning science in informal environments: People, places, and pursuits*. National Research Council. https://doi.org/10.17226/12190

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *FABBS foundation, psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.

Blanco, L., Guerrero, E., Caballero Carrasco, A., Brigido, M., & Mellado, V. (2010). The affective dimension of learning and teaching mathematics and science. In M. P. Caltone (Ed.), *Handbook of lifelong learning developments* (pp. 265–287). Nova Science Publishers.

Campbell, J. L., Quincy, C., Osserman, J., & Pederson, O. K. (2013). Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research*, 42(3), 294–320.

Chung, H., & Gerber, E. (2010). Emotional-storyboarding: A participatory method for emotional designing for children. *Proceedings from the 7th international conference on design & emotion*. Chicago 4–7, October. Chicago: IIT Institute of Design.

Cross, R. M., & Warwick-Booth, L. (2016). Using storyboards in participatory research. *Nurse Researcher*, 23 (3), 8–12.

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

——— , 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

——— , Lehman, B., & Person, N. (2010). Monitoring affect states during effortful problem solving activities. *International Journal on Artificial Intelligence in Education*, 20, 361–389. https://doi.org/10.3233/JAI-2010-012

——— , Person, N., & Lehman, B. (2009). Antecedent-consequent relationships and cyclical patterns between affective states and problem solving outcomes. In *Proceedings of AIED*, 57–64. https://doi.org/10.3233/978-1-60750-028-5-57

DeWitt, J., & Osborne, J. (2010). Recollections of exhibits: Stimulated-recall interviews with primary school children about science centre visits. *International Journal of Science Education*, 32, 1365–1388. https://doi.org/10.1080/09500690903085664

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, 41(10), 1040–1048.

Earl, S. R., Taylor, I. M., Meijen, C., & Passfield, L. (2017). Autonomy and competence frustration in young adolescent classrooms: Different associations with active and passive disengagement. *Learning and Instruction*, 49, 32–40. https://doi.org/10.1016/j.learninstruc.2016.12.001

Ekman, P. (1993). Facial expression and emotion. *American Psychologist*, 48(4), 384–392.

Falk, J. H. (2016). *Identity and the museum visitor experience*. Routledge. https://doi.org/10.4324/9781315427058

——— , & Dierking, L. D. (2000). *Learning from museums: Visitor experiences and the making of meaning*. AltaMira Press.

Festinger, L. (1962). *A theory of cognitive dissonance* (Vol. 2). Stanford University Press.

Forman, J., & Damschroder, L. (2008). Qualitative content analysis. In L. Jacoby, & L. A. Siminoff (Eds.), *Empirical methods for bioethics: A primer*, 1st ed. Elsevier JAI.

Graesser, A., Ozuru, Y., & Sullins, J. (2010). What is a good question? In M. G. McKeown, & L. Kucan (Eds.), *Threads of coherence in research on the development of reading ability* (pp. 112–141). Guilford Press.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Exploratory factor analysis: Multivariate data analysis, 7th Pearson new international ed*. Pearson.

Hein, G. E. (1998). *Learning in the museum*. Routledge.

Immordino-Yang, M. H., Darling-Hammond, L., & Krone, C. (2018). *The brain basis for integrated social, emotional, and academic development: How emotions and social relationships drive learning*. Aspen Institute.

Järvenoja, H., & Järvelä, S. (2009). Emotion control in collaborative learning situations: Do students regulate emotions evoked by social challenges? *British Journal of Educational Psychology*, 79(3), 463–481. https://doi.org/10.1348/000709909X402811

Kapur, M. (2008). Productive failure. *Cognition and Instruction*, 26(3), 379–424.

Lodge, J. M., Kennedy, G., Lockyer, L., Arguel, A., & Pachman, M. (2018). Understanding difficulties and resulting confusion in learning: An integrative review. *Frontiers in Education*, 3(49). https://doi.org/10.3389/feduc.2018.00049

Lyle, J. (2003). Stimulated recall: A report on its use in naturalistic research. *British Educational Research Journal*, 29(6), 861–878. https://doi.org/10.1080/0141192032000137349

Meier, G., & Albrecht, M. H. (2003). The persistence process: Development of a stage model for goal-directed behavior. *Journal of Leadership & Organizational Studies*, 10(2), 43–54. https://doi.org/10.1177/107179190301000205

Nakamura, J., & Csikszentmihalyi, M. (2002). The concept of flow. In C. R. Snyder, & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 89–105). Oxford University Press.

Norris, L., & Tisdale, R. (2017). Developing a toolkit for emotion in museums. *Exhibition*, 36(1), 100–108.

Packer, J., & Ballantyne, R. (2002). Motivational factors and the visitor experience: A comparison of three sites. *Curator: the Museum Journal*, 45(3), 183–198. https://doi.org/10.1111/j.2151-6952.2002.tb00055.x

Paneto, S., Todd, K., Rappolt-Schlichtmann, G., Kipling, B., Marino, K., Malandain, B., & May, S. (2021). *Designing for productive struggle: A research and development guide to creating exhibits that are both challenging and rewarding*. Museum of Science.

Papert, S. (2002, June 24). Hard fun. *Bangor daily news*. Retrieved from https://archive.bangordailynews.com/2002/06/24/how-to-make-writing-hard-fun/.

Paris, S. G. (1997). Situated motivation and informal learning. *Journal of Museum Education*, 22(2–3), 22–27. https://doi.org/10.1080/10598650.1997.11510356

Paris, S. G., & Newman, R. S. (1990). Development aspects of self-regulated learning. *Educational Psychologist*, 25(1), 87–102. https://doi.org/10.1207/s15326985ep2501_7

Pedretti, E., & Navas Iannini, A. M. (2020a). *Controversy in science museums: Re-imagining exhibition spaces and practice*. Routledge.

Pedretti, E., & Navas Iannini, A. M. (2020b). Towards fourth-generation science museums: Changing goals, changing roles. *Canadian Journal of Science, Mathematics and Technology Education*, 20(4), 700–714.

Pekrun, R. (2011). Emotions as drivers of learning and cognitive development. In R. A. Calvo, & S. K. D'Mello (Eds.), *New perspectives on affect and learning technologies* (pp. 23–39). Springer. https://doi.org/10.1007/978-1-4419-9625-1_3

——— (2014). *Emotions and learning. Educational practice series – 24*. International Academy of Education/International Bureau of Education. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.691.9950&rep=rep1&type=pdf

———, Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, 37(2), 91–105. https://doi.org/10.1207/S15326985EP3702_4

———, & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 259–282). New York, NY: Springer. https://doi.org/10.1007/978-1-4614-2018-7_12.

Perry, D. (2012). *What makes learning fun? Principles for the design of intrinsically motivating museum exhibits*. AltaMira Press.

Rappolt-Schlichtmann, G., Evans, M., Reich, C., & Cahill, C. (2017). Core emotion and engagement in informal science learning. *Exhibition*, 36(1), 42–51.

———, Todd, K., & Daley, S. (2020). Universal design for emotion in learning: A practice for the creation of emotionally accessible digital learning experiences. In R. Marquis, & J. Majewski (Eds.), *Inclusive digital interactives: Best practices, innovative experiments, and questions for research* (p. 407). Access Smithsonian. Retrieved from https://access.si.edu/sites/default/files/inclusive-digital-interactives-best-practices-research.pdf

Rittmayer, A. D., & Beier, M. E. (2009). Self-efficacy in STEM. In B. Bogue, & E. Cady (Eds.) *Applying research to practice (ARP) resources*. Retrieved from http://aweonline.org/arp_selfefficacy_overview_122208.pdf.

Simon, R. A., Aulls, M. W., Dedic, H., Hubbard, K., & Hall, N. C. (2015). Exploring student persistence in STEM programs: A motivational model. *Canadian Journal of Education*, 38(1), 27.

Sinatra, G. M., Broughton, S. H., & Lombardi, D. (2014). Emotions in science education. In R. Pekrun, & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 415–436). Routledge/ Taylor & Francis Group.

Staus, N. L., & Falk, J. H. (2017). The role of emotion in informal science learning: Testing an exploratory model. *Mind, Brain, and Education*, 11(2), 45–53. https://doi.org/10.1111/mbe.12139

Thagard, P. R. (2002). The passionate scientist: Emotion in scientific cognition. In P. Carruthers, S. P. Stich, & M. Siegal (Eds.), *The cognitive basis of science* (pp. 235–250). Cambridge University Press.

Truong, K. N., Hayes, G. R., & Abowd, G. D. (2006). Storyboarding: An empirical determination of best practices and effective guidelines. *Proceedings of the 6th conference on designing interactive systems*, 12–21. https://doi.org/10.1145/1142405.1142410

Velayutham, S., Aldridge, J., & Fraser, B. (2011). Development and validation of an instrument to measure students' motivation and self-regulation in science learning. *International Journal of Science Education*, 33 (15), 2159–2179. https://doi.org/10.1080/09500693.2010.541529

Vermeer, H. J., Boekaerts, M., & Seegers, G. (2000). Motivational and gender differences: Sixth-grade students' mathematical problem-solving behavior. *Journal of Educational Psychology*, 92(2), 308–315. https://doi.org/10.1037/0022-0663.92.2.308

Vincent-Ruz, P., & Schunn, C. D. (2018). The nature of science identity and its role as the driver of student choices. *International Journal of STEM Education*, 5(1), 48. https://doi.org/10.1186/s40594-018-0140-5

Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*. Harvard University Press.

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), 1063–1070. https:// doi.org/10.1037/0022-3514.54.6.1063