Designing for Productive Struggle: A Research and Development Guide to Creating Exhibits that are Both Challenging and Rewarding

Written by *Sunewan Paneto, Katie Todd, Rebecca (Becki) Kipling, Gabrielle Schlichtmann, Katharina Marino, Beth Malandain, and Sarah May*

Designed by *Sara Castellucci*

Museum of Science. EdTogether UNIVERSITY of ROCHESTER CAST

Acknowledgments

This project was a collaborative, interdisciplinary effort between two core groups: the Research Team and the Exhibits Team. The contents of this Guide would not have been possible without each team member's knowledge, passion, and insight. We would like to recognize the following organizations and individuals for their contributions to the project:

Thank you to the **Research Team** for their work in collecting, analyzing, and distilling the data gathered for this project. This team was instrumental in diving into the emotions literature, and iterating the methods used to explore productive struggle, which are now described in this Guide.

Museum of Science
Keith Allison
Allison Anderson
Marta Beyer
Clara Cahill
Megan Gregory
Annemarie Gundel
Alex Lussenhop
Elizabeth Kunz Kollmann▪
Sarah May•
Leigh Ann Mesiti
Sunewan Paneto•
Christine Reich▪
Katie Todd•
Owen Weitzman

EdTogether
Amanda Cook
Miriam Evans
Gabrielle Schlichtmann•▪

University of Rochester
Samantha Daley▪
Pamela Zeidan

CAST
José Blackorby▪
Kim Ducharme
Kevin Kent

Thank you to the **Exhibits Team** for their expertise in content development, design, and fabrication of the productive struggle exhibits. The creation of the productive struggle framework seen in this Guide would not have been possible without their creativity, hard work, and knowledge.

Museum of Science
Charlotte Ames
Ann Atwood
Sara Castellucci•
Rebecca (Becki) Kipling•
Chris Brown
Edward Collins
Donna DiBartolomeo
Donald Greene
Savannah Hubbard
Bill Jackson
Beth Malandain•
Katharina Marino•
Ryan Morey
Chelsea Murphy
Kristen Orr
Doug Packer
Alana Parkes
Magdalena Rabidou
Robert "Boob" Rayle
Danielle Covatta Riva
Joe Rivers
Keith Simmons
Jerry Sullivan
Chris Whiting
Ben Wilson

Emily Robertson Voice Over
Emily Robertson

• Authors of the guide
▪ Project PIs

Acknowledgments

And, another thank you to our **project advisors** for their advice and guidance:

Sidney D'Mello, University of Colorado Boulder
Josh Gutwill, Exploratorium
Lauri Halderman, American Museum of Natural History
J. Shipley Newlin, Independent Consultant
Deborah Perry, Selinda Research Associates, Inc.

And finally, a thank you to all the **participants** in our research for making this work possible!

How to cite this Guide:
Paneto, S., Todd, K., Kipling, R., Schlichtmann, G., Marino, K., Malandain, B., May, S., & Castellucci, S. (2021). *Designing for Productive Struggle: A Research and Development Guide to Creating Exhibits that are Both Challenging and Rewarding.* Museum of Science, Boston.

This work was supported by a grant to the Museum of Science, Boston from the National Science Foundation under Award # DRL-1612577. Any opinions, findings, conclusions, or recommendations expressed in this document are those of the authors and do not necessarily reflect the views of the Foundation.

Table of Contents

Part I:
Emotions and Productive Struggle

Introduction

You feel anticipation as you creep forward slowly, inching towards a screen showing a bird sitting on a tree branch above some bushes. Your quick steps cause the bird to quit singing as it turns to look and listen. You experience a momentary unease, a pit in your stomach, and a tightness in your shoulders. You pause to catch your breath. The bird turns away and resumes its song. You step forward quickly again, this time the bird definitely notices you and chirps in alarm. A notification appears on the screen that reads: "You were too fast. The bird's call alerted the animals and they ran away!" You feel confused and surprised. How did this happen? You read the exhibit label and learn that some animals in the wild listen to birds, so if you scare a bird and it makes an alarm call, you may spook other animals too. A prompt suggests you try again. You return to the starting position with new determination and begin to creep forward once more, this time willing your body to be smooth and slow, acting relaxed even though, on the inside, your heart is pounding with anxiety. As you approach the bird you feel more confident and are encouraged as you notice the bushes slowly begin to part. You reach the end of the walkway and the screen reveals a family of deer! You realize how tense you'd been as your shoulders drop and you feel a sense of relief mixed with pride. You can't help but grin. "I did it!" you yell to your friends, and they clap.

Our story

Museums are known for emotional experiences. For years, our field has sought to develop learning experiences that support feelings like curiosity, excitement, wonder, and awe. In 2012, a team of researchers from the Museum of Science, Boston (the Museum) and CAST began an exploratory research study aimed at describing the range of emotions that visitors feel in an exhibition. Through a National Science Foundation-funded project called *Pathways: Emotion and Thinking in Designed Informal Science Environments* (DRL-1222613), we found that visitors experienced deeper engagement when they reported a mix of negative and positive emotions, along with a feeling of overall satisfaction (Rappolt-Schlichtmann et al., 2017).

This finding—that negative emotions could contribute to valuable outcomes for visitors—surprised us and sparked our interest. Our previous approaches had primarily focused on designing experiences where visitors would feel positive emotions. Yet, the *Pathways* study showed that many people experienced negative emotions in our exhibits.

This finding from the Pathways study made us wonder: What if, rather than trying to make everyone feel good, we focused on supporting people to productively persist through their negative emotions to ultimately experience satisfying outcomes?

To explore how design might support people through negative emotions, we applied for and received a follow-up grant to undertake a project titled *Guidelines for Designing Challenging and Rewarding Interactive Science Exhibits* (DRL-1612577). In this design-based research (DBR) project, the Museum, EdTogether, CAST, and the University of Rochester collaboratively investigated how design could support an experience like the one described in the vignette at the beginning of this section–an experience that our team now calls **productive struggle**.

We define productive struggle as an experience with three elements:

1. A learner encounters a disruptive task, phenomena, or idea and shifts into a state of disequilibrium (which might be experienced as emotions like confusion, frustration, surprise, or unease).
2. The learner is supported to persist through disequilibrium using emotional or behavioral resources (e.g., motivation, self-efficacy, problem-solving, trying again).
3. The learner achieves an emotionally productive resolution tied to the source of disequilibrium or a more holistic sense of effortful achievement.

See "What is Productive Struggle?" beginning on page 20 for more details.

Our team has developed several exhibits where learners experience productive struggle, and we have created and refined a framework of design strategies for developing such exhibits. Our research demonstrates that when we design for productive struggle, visitors have experiences they describe as **valuable, engaging, and educational** (Todd et al., 2021). Working extensively with the Universal Design for Learning Framework (udlguidelines.cast.org) in the past, our team had been able to support a wider range of people to perceive, understand, navigate, interact with, and contribute to museum exhibits. However, this project helped us recognize how the Universal Design for Learning approach underattends to emotional factors in explicit ways. We suggest that emotional accessibility in exhibit design should recognize the role of emotions in learning and the value of embracing and designing for variability in visitors' emotional experiences and preferences in educational settings.

▸ Broadly speaking, the goal of emotionally accessible design is to enhance visitors' feelings of belonging, engagement, and motivation for learning.

More specifically, the research described in this book shows that careful design can offer visitors access to complex emotional experiences, like productive struggle, that are critical to STEM learning, but challenging for some learners to initially embrace.

How to use this book

This guide was created with museum professionals–including exhibit developers and designers, researchers, and evaluators–in mind. You do not need to read this guide from cover to cover. Each part of this guide was written to stand alone, so you can skip around to the sections that are most meaningful to you and your work. Whether you hope to design productive struggle exhibits, evaluate them, or think more broadly about emotion in your context, we intend that this guide will provide background information and strategies for exploring research-based practice that attends to emotional factors to benefit learners.

A Brief Summary of Each Section

Part I: Emotions and Productive Struggle

- **Introduction:** A brief overview of the team that developed the Productive Struggle Framework and the purpose of this guide.
- **Why Emotion?:** A summary of what our team has learned about why emotions are important to consider in creating museum experiences.
- **Emotion 101:** A brief overview of emotion science that has informed our work, along with definitions for commonly used terminology and discussion of concepts from the emotions literature.
- **Negative can be Positive:** A reflection on how we came to see negative emotions as valuable in learning.
- **What is Productive Struggle?:** An explanation of our definition of productive struggle, its core components, and how this definition was developed.
- **Is Productive Struggle a New Idea?:** An overview of prior research, including learning theories and frameworks, we drew on in our work on productive struggle.
- **Developing the Productive Struggle Framework: A Design-Based Research Process:** A brief explanation of our design-based-research (DBR) process for this project.

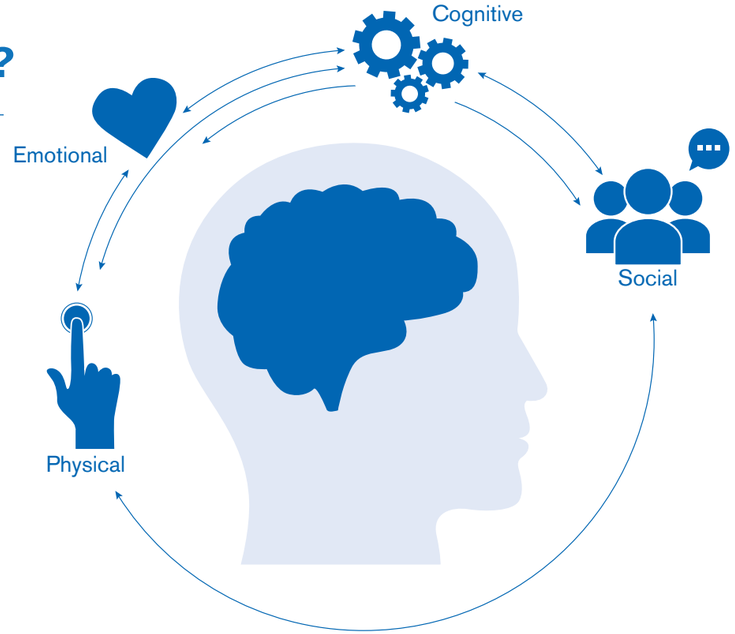
Part II: The Productive Struggle Framework

- **Overview:** A brief intro to our framework and how to use it.
- **The Productive Struggle Framework (Graphic):** A high-level, visual summary of our team's research-based framework, which presents strategies for designing productive struggle exhibits and describes resulting emotional outcomes for visitors.
- **Applying the Productive Struggle Framework:** A detailed explanation of the stages that compose productive struggle experiences and the design strategies that support visitors' movement through those stages.
- **Examples from Research & Practice:** An introduction to this section is followed by a collection of "Case Examples" that describe how we designed three different exhibits to elicit productive struggle, lessons learned, and evaluation findings, along with "Framework Implementation Examples" that outline how designers implemented the framework to elicit productive struggle in each exhibit.

Part III: Measuring Productive Struggle

- **How Do We Measure Productive Struggle?:** An overview of the methods our research team used to study productive struggle.
- **Conclusion:** Team reflections and next steps to consider.
- **FAQ:** Commonly asked questions regarding productive struggle and emotions.
- **Glossary:** Definitions for some common terms used in this guide.
- **Instrument Appendix:** A collection of instruments we used throughout this project, along with instructions, for readers to use.
- **References:** A list of reference materials we used in the development this guide.

Why Emotion?



> "Without emotion, all decisions and outcomes are equal—people would have no preferences, no interests, no motivation, no morality, and no sense of creativity, beauty, or purpose… Emotions are, in essence, the rudder that steers thinking." (Immordino-Yang, 2015, pp. 27-28)

Our work around emotions came about as the Museum of Science, Boston (the Museum) sought to expand its efforts in creating accessible STEM learning experiences. For years, the Museum has been a leader in applying principles from the Universal Design for Learning (UDL) framework to develop educational and accessible experiences for public audiences. This framework, developed by CAST (cast.org) and originally released in 2009, is updated on a regular basis to integrate the latest insights from fields that comprise the learning sciences. In recent years, with advances in affective neuroscience, CAST's UDL framework has shifted to include the role of emotion in accessible practices (Posey, 2018). While popular belief has seen emotion as at odds with rationality, current thinking within the neuro- and psychological sciences sees emotion interwoven into every aspect of our lives—affecting how we perceive, understand the world, relate to others, think, learn, and develop. Emotion, as these fields now see it, is essential for rationality.

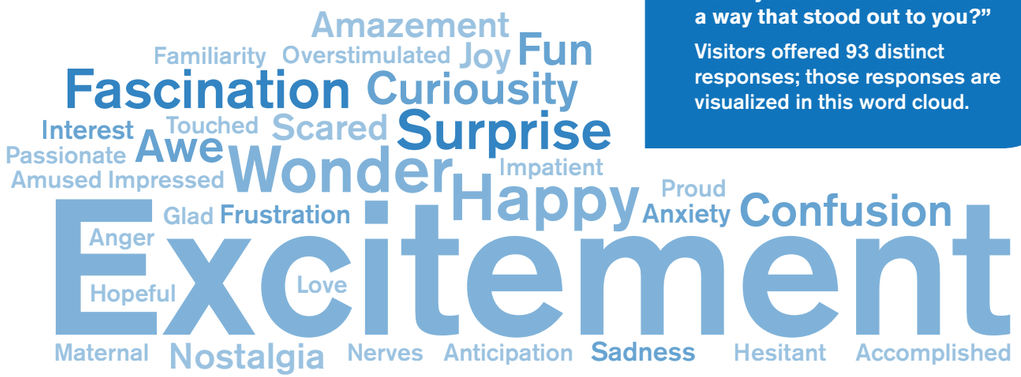
While the study of emotion is a relatively new science, the research literature is vast. To begin applying cutting-edge affective scholarship to the design of informal science learning experiences, it was essential to prioritize interdisciplinary collaboration. Reflecting this strategy, the Productive Struggle Team is composed of researchers, evaluators, developers, and designers who represent a range of backgrounds, experiences, skills, and expertise (including affective sciences, informal learning design and research, and UDL). Our cross-disciplinary team includes staff from EdTogether, CAST, and the University of Rochester, as well as the Museum. While we do not think it is necessary for teams that want to design for productive struggle to have such an array of people directly available to them, it was certainly helpful as our team navigated work across fields that have a limited history of cross-pollination.

Prior to this work, those of us from the museum field often spoke about emotion-relevant concepts without recognizing them in that way. For example, constructs like engagement, interest, identity, and attitudes are recognized as important informal learning design outcomes, but are also intertwined with emotional experiences (Friedman, 2008; National Research Council, 2009). When addressing these outcomes in design, we had often used emotion terms imprecisely and without purposeful connection to the latest findings from the emotion sciences. Compared to other factors, like cognition and behavior, emotions seemed nebulous and unmeasurable.

Yet, in recognizing that it is not possible for visitors to leave any part of themselves behind when they visit a museum, we came to see exploring emotion as critical to our understanding of serving the whole person. Everywhere we looked, we started to see the crucial role of emotion in our work as we crafted novelty, sought relevance, fostered social interaction, and stimulated active thought. We realized that, whether or not we had intentionally designed for emotions, visitors were already experiencing many of them in the Museum, ranging from happiness, awe, and empathy to anxiety, frustration, and fear.

We asked ourselves: What if we thought specifically about emotion for its own sake? What could we gain from setting emotional goals in a museum? How does the diversity of learners' emotional experiences and emotional intelligence influence the museum experience?

We began by trying to measure, describe, and understand visitors' emotional experiences in designed museum spaces. When we asked visitors to describe a time they felt emotion in the Museum, we were amazed at the rich diversity of visitors' feelings, as visualized in this word cloud:



We asked 37 visitor groups (102 total visitors),

"Just thinking about today's visit, could you describe a time when you felt emotion here in a way that stood out to you?"

Visitors offered 93 distinct responses; those responses are visualized in this word cloud.

Our work progressed to thinking about a specific, complex emotion state (productive struggle) that involves a mixed experience of negative emotions, like confusion and frustration, along with positive feelings, like pride and satisfaction. Our research shows that productive struggle is associated with increased dwell times at exhibits and that visitors find productive struggle experiences to be highly engaging, educational, and valuable (Todd et al., 2021). Productive struggle is also an important life skill that museums can foster. We find the relevance to science practice especially valuable for our science museum context; for example, Lin-Siegler et al. (2016) share about the authenticity of the need for scientists to persist through challenges in their article "Even Einstein Struggled."

We hope that others will join in this line of inquiry—not only to design for specific emotional trajectories, like productive struggle, within exhibits, but also to continue exploring and defining what it means to design for emotion, holistically, in informal learning environments. We argue that doing so can help better support visitors' experiences and promote positive engagement in learning. In addition to designing exhibits with learning and experience goals, striving for emotional goals can help us create better museum experiences for more people. The box on the next page shares several reflections about how museums can leverage emotions to improve their work.

Productive struggle beyond this project

The work in this guide reflects insights from our research, which focused on adolescents who interacted with science museum exhibits. However, we would guess that much of what we have learned also applies to younger and older audiences, to learning that happens outside of museums, and to content other than science. We encourage you to think about exploring productive struggle in these other contexts (and we would love to hear what you discover!).

Emotional Accessibility in Informal Learning

Emotionally accessible exhibit design explicitly attends to the ways emotion is intertwined with learning. We encourage exhibit professionals to consider three aspects of emotional accessibility:

Promoting emotional safety

Along with physical and cognitive accessibility, emotion plays a critical role in promoting a sense of welcome and belonging, feelings that are essential to environments where people are physiologically prepared to fully engage. For example, when people are in emotional states of perceived threat, they are less able to process new information (Tyng et al., 2017). Designing for emotional safety can help address some of the most basic aspects of free-choice learning, such as whether visitors choose to approach or avoid certain exhibits and, once there, whether they experience a sense of control, relevance, and comfort.

Designing with emotional intelligence

Affective science has generated a wealth of knowledge about how emotions influence attention, motivation, decision-making, memory, and the way we process information. We can apply this knowledge in museums by designing for emotions that support our goals. For example, research has shown that pleasant, low-energy emotional states can facilitate consensus-building (Brackett, 2019). If a museum sought to foster collaborative decision-making, it might do some matchmaking by designing a communal space that evoked feelings of peaceful relaxation to facilitate this goal. By finding evidence-based matches between research findings and exhibit goals, we can improve our practice.

Strengthening learners' emotional skills

The growing body of literature around social-emotional skills and emotional intelligence suggests that these skills have far-reaching benefits. These range from improved health, to academic performance, to employment outcomes—including the identification of key social-emotional intelligences for STEM (Zeidner et al., 2009). We can design learning experiences that allow visitors to practice, develop, and leverage emotional skills to enhance their outcomes across these domains. For example, our productive struggle project supports learners to work through disequilibrium to achieve a satisfying outcome. This is an authentic scientific practice. STEM professionals routinely regulate their emotions when they confront novel ideas, phenomena, and research outcomes. In this way, building one's capacity for productive struggle may enhance future STEM participation.

Emotion 101

One key component of our early work was to try to answer the questions:
What is the nature of emotion? What aspects of emotion should we attend to?

Several distinct, but related aspects of emotional experience are defined within the affective science research literature:

Appraisal is the way you judge or assess an experience, including whether an experience is: good or bad; relevant or irrelevant to your goals; comfortable or threatening; novel or familiar; within or outside of one's control; and consistent with or opposed to social norms. Appraisals are associated with bodily changes in heart rate, blood pressure, and respiration–changes which substantially inform our emotional experience. In educational settings, learners make ongoing and largely subconscious evaluations in anticipation of or during learning that inform their emotional experience.

Core affect constitutes your basic bodily state measured in two dimensions: 1) feeling pleasant or unpleasant and 2) the level of activation or energy you feel in your body. Core affect can be represented on a four-quadrant grid, as shown on the next page. People's core affect changes on a constant basis in response to the environment and their own experiences. Affective scientists generally agree that core affect is universal to human beings and is present from birth.

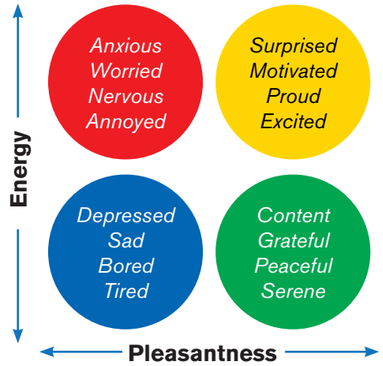
Subjective feeling is the way you metacognitively conceptualize your overall experience of core affect, assigning it meaning by labeling bodily sensations with familiar emotion terms like happiness, sadness, rage, pride, relief, and others.

Taken together, appraisals and core affect constitute a kind of real-time, sometimes unconscious barometer or description of our relationship to our environment, which is strongly associated with motivated behaviors like whether someone would approach or avoid an exhibit. Subjective feeling is the way we then emotionally label these experiences. Even though everyone makes appraisals, experiences core affect, and processes it as subjective feeling, the content of these judgments and interpretations vary between individuals.

Dissecting an Emotional Experience

I walk by an enclosure with an alligator. The creature is behind glass, but reptiles make me uncomfortable (appraisal). My heart rate speeds up. My body feels a little bit bad–a kind of general unpleasantness–and slightly activated (core affect). I am anxious (subjective feeling).

For core affect, each person inhabits their own affective "home base" or default tendencies. Some may frequently shift from positive to negative and experience large fluctuations in energy levels, while others may tend to stay in one general "zone." Meanwhile, subjective feeling is culturally sensitive (e.g., cultures vary in emotion vocabulary, concepts, and awareness of core affect) and observable in many ways, including social signals, vocal patterns, language, gestures, and facial expressions. Thus how people appraise a situation and interpret these changes in core affect depends on their personal expectations, as well as cultural and environmental factors.



Core affect can be measured along two dimensions: unpleasant to pleasant, and low energy to high energy. Examples of subjective feelings are illustrated in the quadrants above.

Social and emotional intelligence (SEI) is our capacity to understand, use, and manage emotions. SEI encompasses a variety of different skill sets. Some examples are:

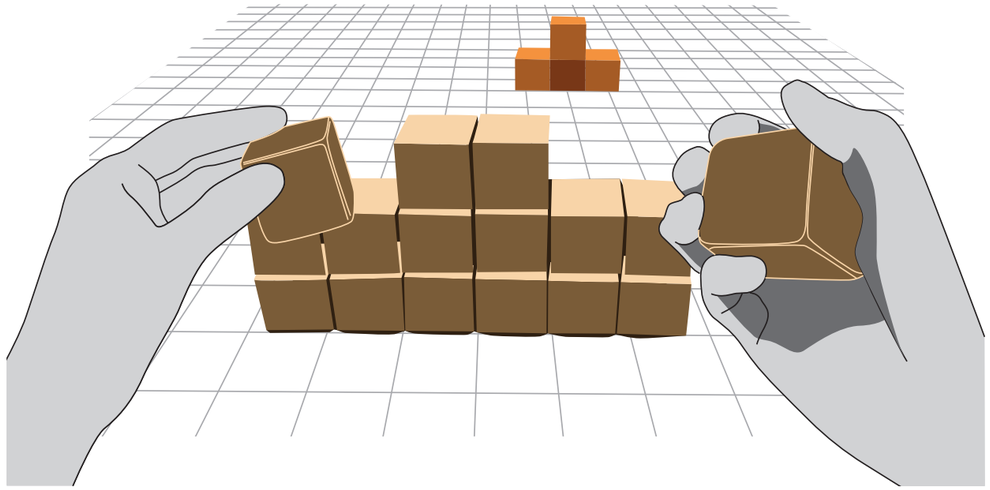
- Self-management (controlling impulses, managing stress and setting goals);
- Social awareness (perspective taking and empathy);
- Relationship skills (cooperating, collaborating, and active listening); and
- Decision making (analyzing and solving problems, reflecting, and ethical responsibility).

Our team has come to view SEI as a critical aspect to address in exhibit design, as visitors must draw on their SEI skills as they work to positively and productively navigate museum spaces. However, we also know that learners vary in their SEI skills. Through exhibit design, visitors can be supported to further develop these skills during their museum experiences. The Productive Struggle Framework provides exhibit design strategies that encourage visitors with varying emotional skills to experience struggle and feel supported while practicing emotional skills essential for STEM learning, including emotional self-regulation and persisting through a challenge.

These critical facets of emotional experience (appraisal, core affect, subjective feeling, and SEI) are the foundation of human motivation. In designing for emotion, museums can provide for the overall wellbeing of our visitors as they: 1) make decisions about their experiences, 2) regulate emotions in the face of meaningful challenges, and 3) adapt and grow in response to museum experiences.

Here, we present an example of how these facets play out for visitors as they experience an exhibit:

Imagine that you are spending some time at the Museum, and approach Math Moves!*, an exhibition focused on ratio and proportion. You decide to try the "Scaling Shapes" exhibit component, where you practice doubling the size of objects in three dimensions–height, length, and width. You begin by using wooden blocks to double one of the model shapes on the table. When comparing your work to the model you realize something is wrong, but you are not sure what. You are experiencing an impasse or mismatch between incoming information and your prior knowledge about what you know about doubling.*



Emotionally, you feel a little bit uncomfortable. Since noticing the impasse (or the error in your approach), your body feels activated and a little bit unpleasant. Interpreting these feelings, you turn to your friend and say, "I'm confused. I'm not sure where I went wrong, but this definitely doesn't match the model." This feeling has cognitive consequences. You begin deliberating with your friend to resolve this feeling and find a solution. Reexamining your work against the model, you look to the label for hints and other resources. You realize that you were only doubling in one dimension–height–when you also needed to address length and width. "Oh! I see what to do!" Your heart races. Your body is activated. You feel very positive, expressing a wide smile, and experiencing joy as you and your friend work quickly to finish doubling the model. Looking at your work, you are satisfied and feel some pride in solving the problem.

This vignette illustrates what we have learned: when museum exhibits are emotionally accessible and supportive, visitors are better able to have engaging experiences, with their whole selves, in adaptive and constructive ways.

With appraisal, core affect, subjective feeling, and SEI in mind, how would an emotion scientist make sense of this scenario?

First, let's consider **appraisal** and **core affect**. During your experience, you are engaged in a kind of subconscious evaluation of your environment. As you progress through the interaction, your assumptions about what would happen and what to do did not match the feedback you received; there was a mismatch between your expectations and this experience. As a result, you feel negative and somewhat activated, and in this way your subconscious appraisal of the situation as novel and uncertain has informed your core affective experience.

As you explored the activity, worked to process information, and problem-solved, you experienced a **subjective feeling**: confusion. Core affect (in this case negative and somewhat active) is one of multiple factors that inform the emergence of subjective feelings and the conscious experience of emotions such as confusion, happiness, anger, or frustration. You recognize confusion because you have felt it before when working to learn something new.

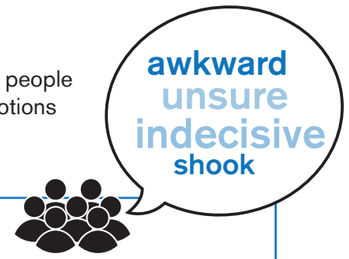
And, finally, what about **SEI**? You felt negative and a little activated at the exhibit. Subjectively, you experienced confusion, but how might you respond? You may stop the activity immediately, overwhelmed with the discomfort of your emotion. Or, you might take a deep breath, consider your situation, and recognize such discomfort as a signal that you have an opportunity to learn something new. With this perspective, you look for support–possibly in your companion or from the exhibit itself. You persist. You resolve the impasse and experience pleasure at your achievement.

This entire emotional process of confusion resolution requires SEI: you stop, think, engage in careful deliberation, problem solve, and revise your existing mental models. In the scenario above, you leverage your social and emotional intelligence, including self-awareness and self-regulation. The choice to engage deeply in a desirable difficulty resulted in a fulfilling emotional experience–and perhaps a greater depth of processing around a new idea and a more durable memory of the experience.

Negative can be Positive

For a long time, the Museum of Science had aspired to create emotional experiences, such as sparking curiosity, inspiration, or awe. But many of us were reluctant to stray beyond wanting "happy visitors." We want people to return to the Museum. Would we not want visitors to have a good time and form positive memories, so hopefully they return to the Museum again and again? Of course there is plenty of space for positive emotions, but when the Museum began working with emotion researchers at CAST, EdTogether, and the University of Rochester, we learned that people were experiencing negative emotions in the Museum, and those emotions were actually associated with positive outcomes!

Could exhibits that embrace emotional variability and support a range of emotions (including negative ones) be pathways to more accessible, equitable learning experiences?



What do we mean by negative emotions?
Affective scientists often categorize emotion words as being pleasant (e.g., happy, proud) or unpleasant (e.g., angry, sad). However, as described below, emotional constructs can vary by culture. This is especially true of complex emotions that we frequently see in museums—like confusion, awe, or surprise—that are prone to have differential perceptions of positivity or negativity based on the context and intensity. In our initial work, we thought about "struggle" with a fairly traditional lens of how affective scientists defined negative emotions, but it soon became clear to us that each person's sense of struggle was different. This led us to expand our thinking about "negative" emotions to include visitors' own interpretations of what felt emotionally unsettling and negative for them, rather than relying solely on how the research literature defined certain emotion terms.

The more we looked into research around negative emotions, the more we found reasons to challenge our bias against them. For example, we began to question how much our discomfort with negative emotions was culturally situated. Research shows there are cultural differences between people's emotional preferences: European Americans tend to value active, positive states like excitement more than Hong Kong Chinese, who tend to value less active, negative states like sadness (Tsai et al., 2006). Individuals within cultures have varied emotional experience and preferences as well. In the United States, younger people tend to strive for more active emotional states like excitement, while older individuals seek out less active states like relaxation and calm (Brackett, 2019). Thus, as a project team that is largely young and white, we began to reflect on how our preference for positive emotions may be a means of perpetuating dominant cultural norms through our work. By designing museum exhibits that foster primarily active and positive feelings, we may be designing spaces that are less emotionally comfortable for diverse audiences.

We also found numerous studies demonstrating the importance of negative emotions for learning (D'Mello et al., 2014; Linn et al., 2010; VanLehn et al., 2003). Although most of this research focused on formal education settings, one study that focused on informal science contexts had complementary results: Staus and Falk (2017) showed that active, negative states (such as nervousness or frustration) contributed to enhanced learning outcomes across a variety of informal learning experiences. These studies suggest that being invested in engaging with negative emotions results in a more memorable learning experience.

In addition, prior research at the Museum showed us that our visitors were already experiencing negative emotions in museums—whether or not we had designed with these emotions in mind (Rappolt-Schlichtmann et al., 2017). Instead of trying to design these emotions away, we can ensure that we are supporting our visitors to moderate their negative emotions and engage with scaffolded negative emotions in ways that feel safe, productive, and less overwhelming than they may encounter in other environments (such as experiencing shame or high anxiety in a classroom). In doing so, we can help visitors build their real-world emotional skills and make their museum visits more meaningful. Later on in the Guide, we outline how our framework can be used to design exhibits that support visitors through these negative emotions. Designing for negative emotions may not only deepen overall engagement and learning but might help museums address difficult or polarizing content in an emotionally accessible way.

Three Reasons to Embrace Negative Emotions

1. Different cultures (and individuals) have varied emotional preferences, so attending to negative emotions can make museums more inclusive.
2. A growing body of research shows that negative emotions can enhance attention, memory, and learning.
3. Visitors already experience—and value—negative emotions that arise from challenging experiences at exhibits.

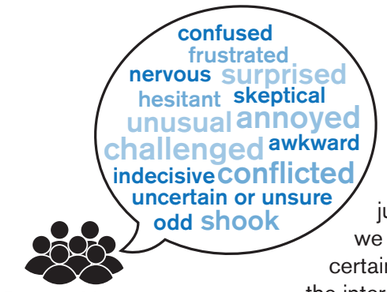
We can reach more people more powerfully when we support a wide range of emotional experiences, including those generally categorized as negative.

What is Productive Struggle?

When we began our project, one of our first tasks was to figure out what we meant by productive struggle. This was an ongoing process of iterating on our definition as we developed and tested exhibits. We started with an understanding that the "struggle" part of our definition included a fairly narrow set of negative emotions such as confusion and frustration. Yet, our preliminary data collection showed that people used many more words to describe their experiences. We began to be concerned that, if our definition was too narrow, we were discounting real instances of struggle when learners used different language to describe it. Given research showing that people have variation in their abilities to articulate certain emotions (Barrett, 2017; Brackett, 2019), we felt that the most inclusive response was to broaden our definition.

Struggle budget

Being in an emotional state of disequilibrium can be taxing on our bodies. Accordingly, using welcoming and accessible design strategies is important to minimize the struggle that visitors experience from usability issues, helping them focus on the challenges we purposefully design.



Disequilibrium

Instead of looking only for confusion and frustration, we began to talk about **disequilibrium**, which we thought of as a state of being emotionally out of balance. In our research, visitors described this imbalance with words like angry, annoyed, awkward, challenged, conflicted, confused, frustrated, hesitant, indecisive, nervous, odd, "shook," skeptical, surprised, uncertain or unsure, and unusual. As we began our prototyping, we did not want just any experience of disequilibrium from our exhibits. Rather, we wanted disequilibrium to be experienced in response to certain design choices we made, and not, for example, because the interactive seemed broken. We wanted visitors to be able to focus their attention on the challenge we had designed, rather than wasting their limited energy on figuring out what to do or how to use the exhibit. We adjusted our design framework (see *The Productive Struggle Design Framework*) and our definition to account for this distinction, noting that disequilibrium should arise from an intended, challenging task within the exhibit, and that a productive struggle exhibit must have inviting and accessible design to limit unnecessary challenges.

So how does struggle become productive? In a free-choice learning environment, visitors can choose to leave whenever they encounter a challenge. To make disequilibrium productive, it is necessary for people to choose to stay at the exhibit and persevere. While our definition of disequilibrium relies heavily on emotional experience, we think of another aspect of our definition, persistence, as more behavioral than emotional. The primary evidence of **persistence** is visitors' observable actions that demonstrate intentional exhibit use and pursuit of a goal, although we did still ask about whether participants had related emotions such as feeling motivated, focused, determined, and persistent (see *How Do We Measure Productive Struggle?* for more details).

Persistence

Productivity

Beyond persistence, we wanted our participants to experience **productivity**. Similar to disequilibrium, we began with a fairly narrow definition of productivity and eventually broadened it. Our initial conversations about productivity focused on participants achieving a goal, which we initially thought about as ones defined by exhibit professionals. However, we soon recognized that participants were defining their own goals, which often complemented what the exhibit professionals intended. Others were setting and achieving goals that were completely irrelevant to the original aim of the exhibit. We also had participants tell us that they achieved a goal but did not seem to experience productivity at an emotional level. This led us to pivot our thinking. We wondered: if disequilibrium is an emotional sense of imbalance, what if productivity were a return to balance? Some of our participants described this resolution with traditionally positive emotion words like happy, proud, or satisfied, while others talked about an easing of their prior tension, such as relief or feeling better.

After many revisions, our team now defines productive struggle as an experience with three elements:

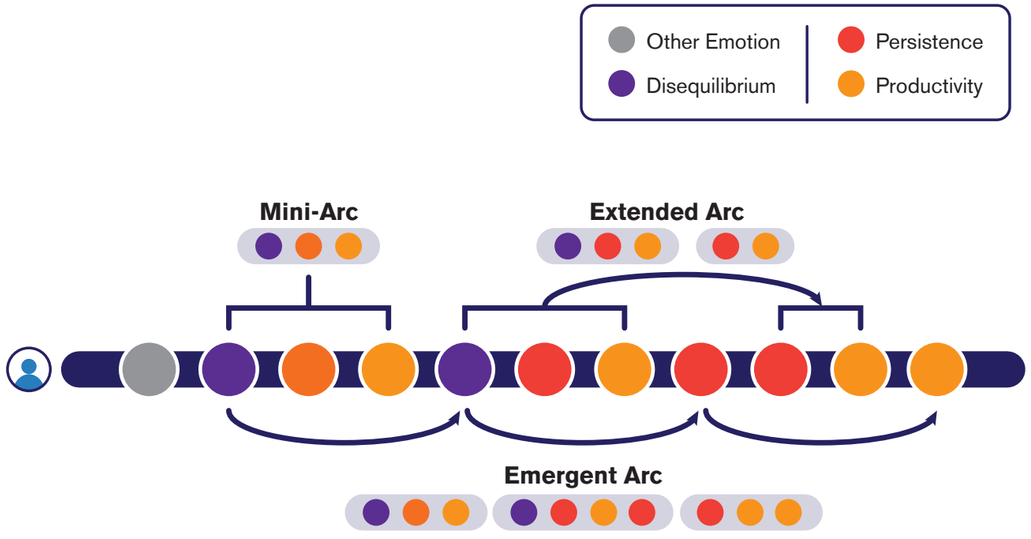
Like a story with an arc, productive struggle has a beginning (disequilibrium), middle (persistence), and end (productivity). However, there are many versions of the productive struggle story. Sometimes, the three part arc occurs very quickly, in a matter of seconds. In other instances, it might be a longer experience that extends over minutes.

We often found that the arc was embedded within a larger exhibit experience: some visitors were at an exhibit for several minutes before encountering disequilibrium, and some remained at an exhibit for a while after experiencing productivity. Our participants showed that productive struggle is not always perfectly linear either: learners often engaged in mini-cycles of productive struggle–they experienced productivity and then chose to re-enter their state of disequilibrium and have a new arc of productive struggle, perhaps digging deeper into a topic or reapplying a skill in a more challenging context (May et al., in press). We see this as encouraging evidence that productive struggle can be a safe and rewarding way to practice persisting through disequilibrium, a skill we hope learners will be able to generalize beyond the Museum's walls.

A Note on Biometric Data

In our Pathways study, evidence from biometric data set us on a course to exploring productive struggle. While we continued to collect biometric data in our productive struggle studies, as detailed in the *How do you Measure Productive Struggle?* section, ultimately, we concluded that visitors' self reports provided the best evidence of visitors' productive struggle experiences.

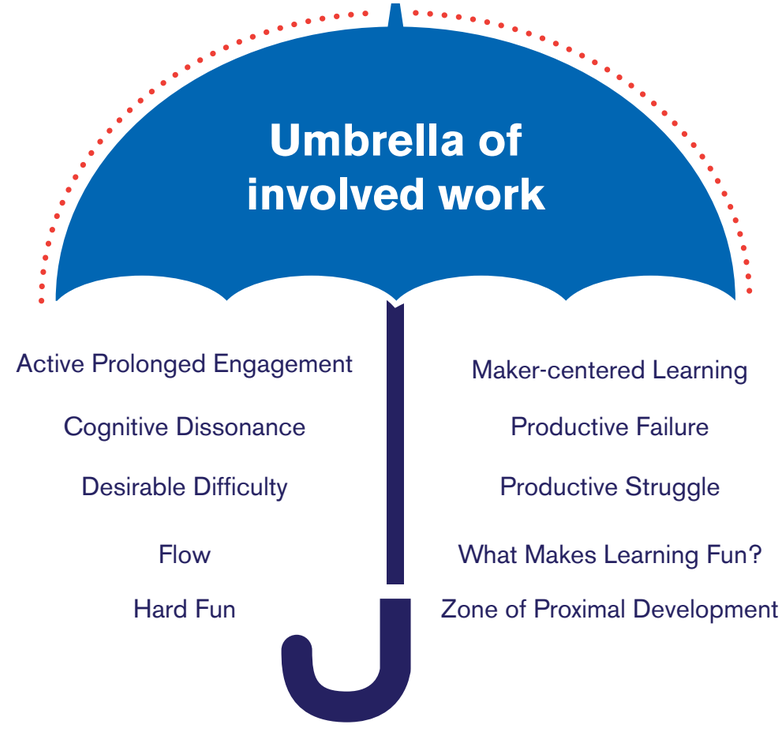
Different Types of Productive Struggle Arcs



Is Productive Struggle a New Idea?

Our work on productive struggle draws on prior research and existing learning theories such as productive failure (Warshauer, 2014, 2015), hard fun (Papert, 2002), desirable difficulty (Bjork & Bjork, 2011), cognitive dissonance (Festinger, 1962), the zone of proximal development (Vygotsky, 1980), and flow (Nakamura & Csikszentmihalyi, 2002). Frameworks for design in informal learning environments have also shaped our thinking, including Active Prolonged Engagement (Humphrey & Gutwill, 2017), What Makes Learning Fun? (Perry, 2012), and efforts from the maker movement (e.g., Bevan et al., 2014; Clapp et al., 2016; Martin, 2015). Researchers in the formal education realm have used the term "productive struggle" before (see Warshauer, 2014; Warshauer, 2015; Granberg, 2016), and have described complementary ideas like discrepant events (e.g., González Espada, 2010; Longfield, 2009; Lynch et al., 2018). Our team sees many relationships between these ideas and, taken together, we refer to these closely related concepts as "involved work." On the following pages, we briefly summarize a selection of these related learning concepts and frameworks, how we have learned from them, and how we think productive struggle is different.

What we think sets our work on productive struggle apart is the focus on ***emotion*** as a central aspect intertwined with cognitive factors of learning.



Active Prolonged Engagement (APE):

Humphrey and Gutwill (2017) define APE as experiences where visitors lead their own learning, have extended dwell times, and show variety in their interactions with an exhibit. Humphrey's and Gutwill's research has had a strong influence on our work, especially in thinking about research design and goals for engagement. Like productive struggle, APE is an example of how exhibits can be designed to foster deeper experiences that focus on more than content learning. In fact, a productive struggle can occur during an APE encounter. However, not all APE experiences involve disequilibrium and not all productive struggles meet the APE criteria for actively visitor-led interactions or variety of engagement.

Cognitive dissonance:

Cognitive dissonance arises from our natural desire to resolve inconsistencies among aspects of "knowledge, opinion, or belief" in order to restore a sense of balance (Festinger, 1962, p. 3). The idea of cognitive dissonance was central to the formulation of our concept of disequilibrium. Like in cognitive dissonance, we recognize that the negative experience of a struggle can motivate learners to change their situation in order to reduce their disequilibrium. Some of our research participants even described their struggles as being cognitively dissonant. In our productive struggle work, though, we took on a broader definition that would also allow for physical, social, and emotional experiences of struggle alongside cognition.

Desirable difficulty:

Desirable difficulties are features of learning experiences that make our brains process information in ways that are more memorable (Bjork & Bjork, 2011). Although we do not focus exclusively on wanting to make visitors better remember the content that we share, our productive struggle framework draws heavily on this idea of intentionally designing difficult learning experiences. We have applied some of the strategies for supporting desirable difficulty and found them to be effective in designing for productive struggle as well–for example, changing the learning conditions over time and giving people minimal information up front with interspersed hints as needed, rather than a big block of information up front. Other strategies from desirable difficulty may be less appropriate for museums, such as giving tests and spacing out learning opportunities over time (Bjork & Bjork, 2011).

Flow:

The concept of flow involves a "complete absorption in what one does" that occurs in intrinsically motivating activities that are at appropriate levels of difficulty (Nakamura & Csikszentmihalyi, 2002, p. 89). Flow theory has informed our work on productive struggle and we have tried to follow its example of deeply connecting learning with affective experiences. Like productive struggle, flow also describes complex affective states that cannot always be pinned down to a simple emotion word. While flow has been a valuable framework for us, we see key differences between it and productive struggle. Whereas productive struggle is characterized by disequilibrium (when a learner is emotionally off balance), "entering flow depends on establishing a balance" (Nakamura & Csikszentmihalyi, 2002, p. 90).

Hard fun:

Hard fun is when a learner chooses to engage in a challenging activity and finds the difficulty enjoyable. Papert (2002) identified hard fun when he noticed that gamers were drawn to games because they were seeking a challenge. We have drawn on this work as a useful example of how a theory can intertwine difficulty and emotion in informal learning experiences. Although we see potential overlap between hard fun and productive struggle for some learners, hard fun is characterized by people purposefully seeking out a challenge and enjoying it in the moment, whereas productive struggle is defined more sequentially, as disequilibrium followed by feelings of pride or satisfaction. We made this distinction in an effort to support productive struggle among a broader audience, including those who might not initially be drawn to a challenge, or who might not fully enjoy it in the moment but nonetheless find it meaningful and emotionally rewarding afterwards.

Maker-centered learning:

Making, design, engineering, and tinkering initiatives have many overlaps with the constructs described under our umbrella of involved work. For example, the Agency by Design project out of Harvard's Project Zero has developed a framework illustrating how sensitivity to design can support maker empowerment, and explores the productive role of failure and struggle in maker-centered learning (Clapp et al., 2016). Others have described the promises of the maker movement and its potential to support learners to engage meaningfully with challenge (Martin, 2015), including in the context of informal science learning environments (Bevan et al., 2014).

Productive failure:

The concept of productive failure arose from a study that found removing scaffolding from a learning activity led to short-term failures but improved long-term understanding (Kapur, 2008). Our work is grateful for this research, which has paved the way in the learning sciences for challenging common ideas about what types of learning experiences can be valuable. We have also made it easy for visitors to have short-term failures in all of our productive struggle exhibits. One key difference between productive failure and productive struggle, however, is that productive failure literature says feelings of frustration should actively be avoided (Kapur & Bielaczyc, 2012), whereas productive struggle embraces these feelings.

Productive struggle (formal mathematics education):

Before our use of the term, "productive struggle" had been picked up in the formal mathematics education realm to describe intentional efforts to design classroom activities that support learners through struggle toward a productive resolution (Warshauer, 2014). Strategies developed to support students through productive struggle resonate with our own work, but are geared toward formal educators, prompting them to question, encourage, give time for, and acknowledge student experiences of productive struggle (Warshauer, 2015). While the framing is similar to our work, we found these approaches emphasized the cognitive–rather than emotional–aspects of productive struggle. Our definition and design framework are unique in that they address the affective components of productive struggle explicitly, and focus on informal science education contexts.

What Makes Learning Fun?:

In her book, *What Makes Learning Fun?* Deborah Perry (2012) describes developing a framework that integrates three perspectives on learning in informal education settings, charecterizing learning in terms of:

1. Motivations: In order to make learning fun, satisfying, and successful for visitors, designers must address six visitor motivations related to communication, curiosity, confidence, challenge, control, and play.
2. Engagements: Visitor learning occurs as processes of social, intellectual, emotional, and physical engagement.
3. Outcomes: Learning also occurs as outcomes or products addressing visitor meaning-making, attitudes and actions, identity, and skills.

Altogether, this framework integrates a broad range of cognitive, emotional, and behavioral aspects of visitor engagement into a set of design principles. Our current work builds from this, but takes a more focused approach by unpacking the experience of productive struggle specifically (rather than visitor learning or engagement generally).

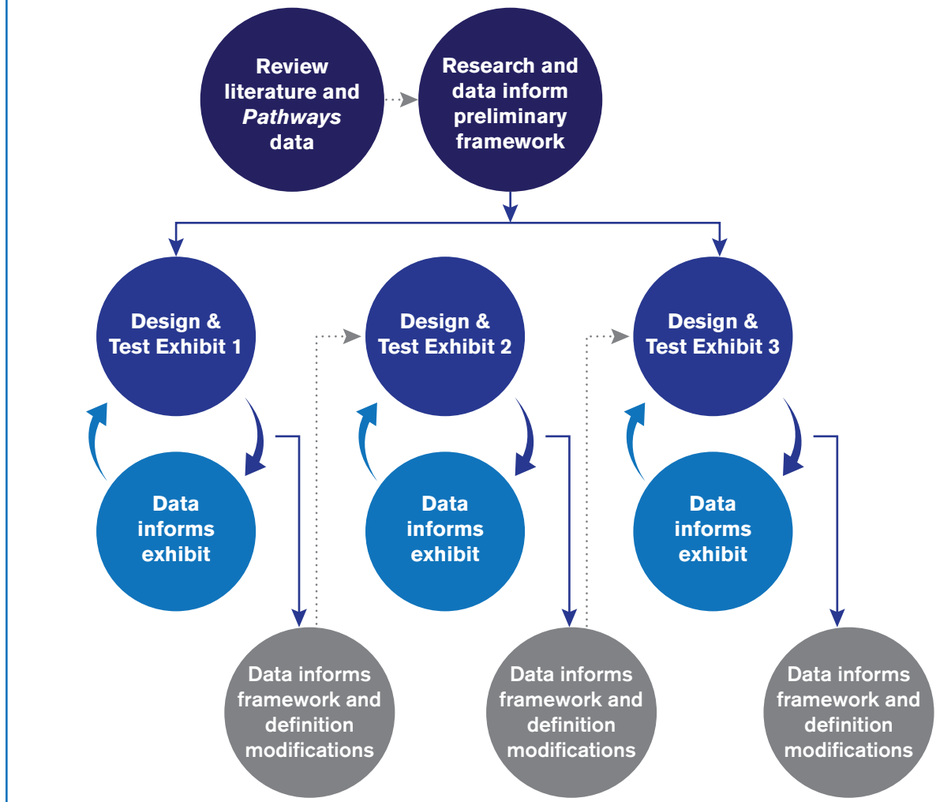
Zone of Proximal Development (ZPD):

Vygotsky's (1980) ZPD is the space between what one is currently able to do and what one can achieve with assistance from a more knowledgeable person. In common use, ZPD has come to mean a "goldilocks" level of learning where scaffolding assists a learner with a task that is neither too hard nor too easy. We have frequently referred to this concept when thinking about our free-choice learning environment: we want our exhibits to be hard, but visitors will leave if something is too hard or there's not enough scaffolding. If we think about the differences between productive struggle and ZPD, we would posit that people likely experience productive struggle in a subset of ZPD learning opportunities; you don't necessarily feel the emotional arc of productive struggle when in the ZPD, but if you do experience productive struggle, you are most likely in the ZPD.

Developing the Productive Struggle Framework: A Design-Based Research Process

This project aimed to develop practical guidelines that define productive struggle and describe how to design museum exhibits that encourage the experience of it. We did this work through a design-based research (DBR) process with collaboration between researchers and exhibit professionals. DBR is typically utilized in formal education as a way "to carry out formative research to test and refine educational designs based on principles derived from prior research" (Collins et al., 2004). Our research team began by reviewing scholarly literature and the findings from the previous *Pathways* study in order to develop a preliminary definition for productive struggle and a framework of design strategies that the project team hypothesized would support it. We then used that preliminary framework to inform our initial exhibit prototypes, which we cyclically tested and revised alongside our definition of productive struggle. This cyclical and iterative process continued as we developed three exhibits to purposefully elicit this type of experience.

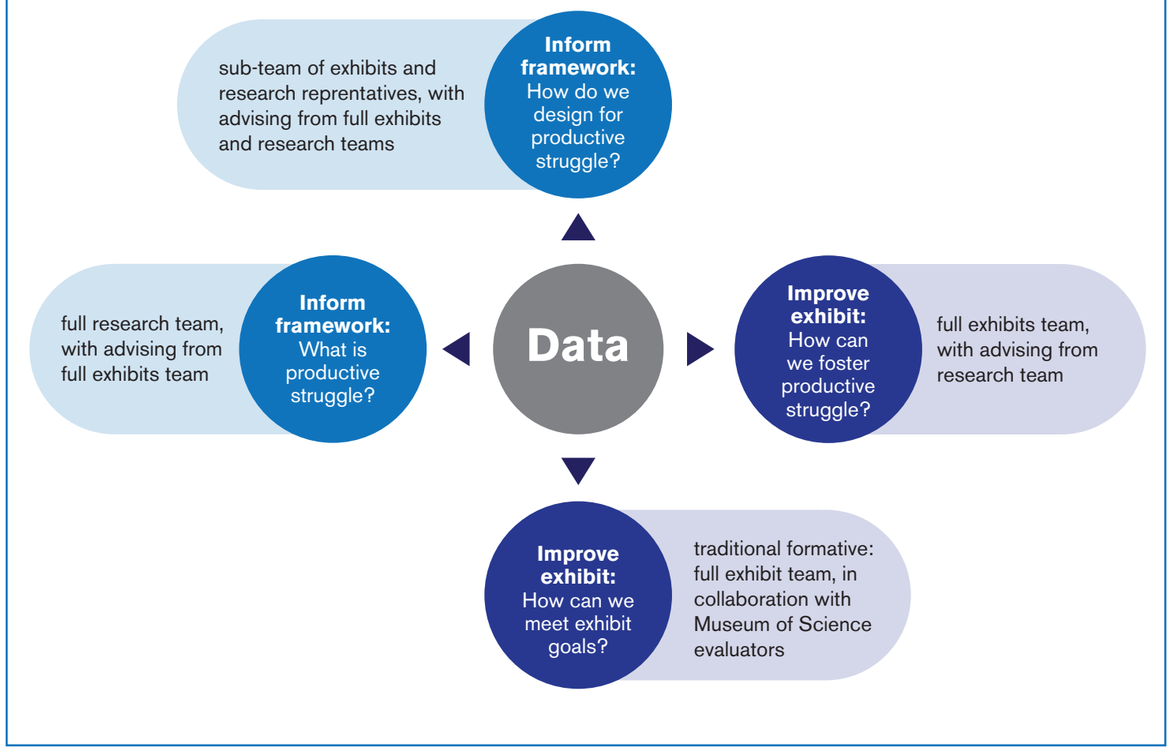
The DBR process that our team used is illustrated in this diagram:



In addition to informing the design framework and definition, the data gathered from each test informed general exhibit improvements. These tests ranged from rapid, small-scale sessions (similar to traditional formative evaluation processes), to larger experimental designs that gathered more conclusive evidence on specific aspects of our framework. In total, we collected data from 455 participants over the course of this project. As a result of this DBR process, everything within the framework is backed by data demonstrating how each design strategy can support productive struggle experiences for youth ages of 10-17.

In general, **DBR can be a messy process because educational contexts are complex**, and it is rarely possible to manipulate single variables like in a laboratory. DBR was particularly challenging in this project because it sometimes felt circular to be simultaneously refining both the definition of productive struggle and the design strategies for supporting it. From the work done in the Pathways study, we were able to establish our definition of productive struggle earlier than the elements of the framework. This helped guide the types of supports we explored and tested. Through multiple testing cycles and resulting adjustments to the framework, we also gradually tweaked and refined our definition throughout the entire project.

In DBR, data informs multiple improvement processes, as illustrated in this diagram:



It is important to note that, while we were sometimes able to isolate and compare single design strategies with one another to see which best elicited productive struggle, systematic, one-by-one testing of all design features was not feasible or practical within the scope of this project. Within these limitations, we chose to study and generate preliminary evidence for a large number of strategies rather than gather conclusive evidence about only a few design approaches. As such, this framework should not be seen as a comprehensive list of all strategies that could support productive struggle. Instead, we present a list of strategies for which we have gathered evidence linked to their support of productive struggle in our exhibits.

The *Examples from Research & Practice* section of this Guide provides additional details about our exhibit design process, the data we used to inform that process, and information about certain strategies we investigated that were not effective in supporting productive struggle. We encourage others to build on this work by testing additional strategies and sharing what you learn!

Part II:
The Productive Struggle Framework

Overview

Our design framework shares strategies for developing museum exhibits where visitors experience productive struggle. As described in the previous section, we tested many strategies for supporting productive struggle–some effective and some not. The final framework, as described in this chapter, represents the culmination of our design-based research processes. Our research showed that each element of this framework was effective in supporting productive struggle at the exhibits we created. It is important to note that, although we tested many strategies, this framework is not comprehensive of every way that design might support productive struggle. Professionals applying this framework should take creative license to experiment with other ideas, and our team would love to hear about what you learn!

Exhibit design is complex. No single design strategy elicited productive struggle in isolation, neither did we use every strategy at any one exhibit. The process of designing for productive struggle is like selecting courses from a menu to assemble a hearty meal. Whether applying the strategies we have tested in a new context or exploring with new ideas, the framework is designed to be used in a data-driven process of developing a prototype, assessing its effectiveness through user testing, and iterating to improve it. The *How do we Measure Productive Struggle?* section details how we carried out this data-driven process.

The framework describes three stages of a visitor's exhibit experience that practitioners should consider when designing for productive struggle: **Invite, Disrupt,** and **Support**. The following pages describe each of these stages, including a brief overview of the goal for that stage, options for design strategies, and guidance about how to apply the design strategies (e.g., do at least one strategy, as many as possible, or do all the strategies).

To help organize all of the strategies and options we have identified in designing for productive struggle, on the next two pages we offer a visual summary of our framework as a quick reference.

While the majority of the design framework focuses on what museum professionals should do when designing exhibits, the bottom of the design framework describes the outcomes of such design in terms of visitor experience, especially with regards to emotional states.

- **Inviting design** contributes to feelings of welcome, such as comfort, safety, and belonging.
- **Disruptive design** leads to disequilibrium, which feels like emotions such as confusion, frustration, and surprise.
- **Supportive design** facilitates both persistence–demonstrated behaviorally and experienced as feelings like focus, motivation, and determination–and productivity, which can feel like satisfaction, happiness, and pride.

Ultimately, designing for productive struggle contributes to experiences that (overall) visitors find engaging, valuable, and educational, and which they describe as feeling like doing science.

The Productive Struggle Framework (Graphic)

Invite

Use clear design to welcome all intended learners to the activity

Do all of these:

Minimize barriers to entry:
- Provide easy orientation
- Demonstrate clear objectives
- Allow visitors to preview
- Make it obvious how to reset or continue
- Present a compelling task
- Maximize relevance, value, and authenticity

Prioritize accessible design for all:
- Design physically inclusive interactions
- Incorporate multisensory features
- Avoid reliance on pre-existing skills and specialized knowledge
- Provide for varied emotional preferences and skills

Disrupt

Facilitate disequilibrium by challenging norms or expectations

Do at least one:

Craft novelty
- Challenge expectations
- Embed surprising phenomena, experiences, or events
- Include unfamiliar information

Leverage uncertainty
- Limit available information
- Force decision-making
- Challenge fine or gross motor skills

Introduce social unease
- Invite competition
- Break social norms
- Embrace interpersonal differences
- Offer a performative element

Support

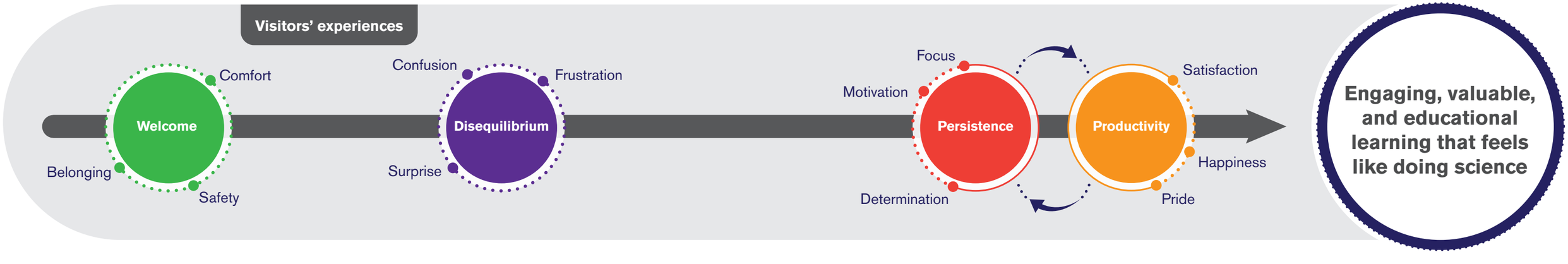
Provide options for persisting through disequilibrium and feeling productive.

Do as many as possible:

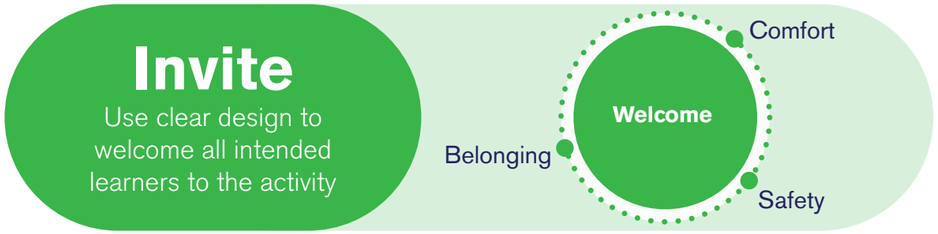
Offer feedback
- Indicate progress or success
- Include mini-wins (throughout) and final rewards
- Integrate hints and scaffolding
- Encourage trying again

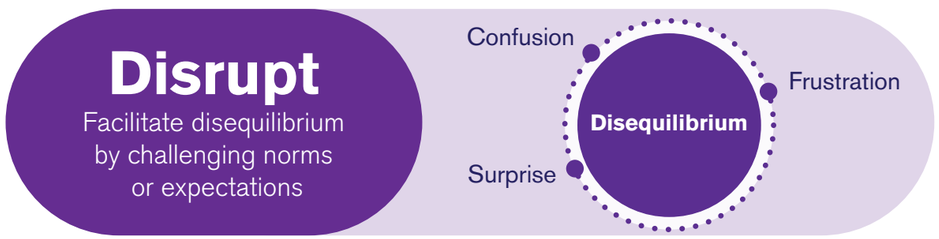
Give choices
- Offer more than one level of challenge
- Create pathways for social and solo interaction
- Design for multiple goals
- Allow repeated attempts
- Include the option to do less/more

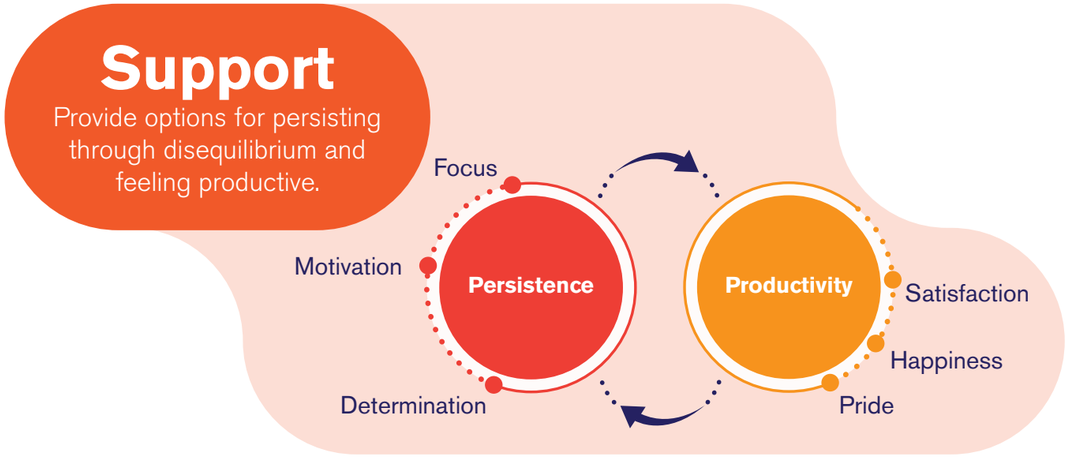
Support self-regulation
- Acknowledge the challenge
- Normalize feelings of disequilibrium
- Invite reflection on disequilibrium



Applying the Productive Struggle Framework







Invite

This stage of the framework focuses on two approaches for inviting visitors to engage with a productive struggle exhibit: **minimize barriers to entry** and **prioritize accessible design for all**.

At first, our design framework began with sparking disequilibrium. But, as visitors used our prototypes, it became important to differentiate between intended and unintended struggle. We did not want visitors to struggle because the exhibit was inaccessible to them or broken. Emotionally, intended and unintended struggle looked similar, but imagine that a visitor is struggling with an aspect of an exhibit and they become frustrated or confused and search for help. If their frustration or confusion is rooted in the usability of the exhibit or factors that are extraneous to the exhibit's intended design, the visitor's disequilibrium could detract from their learning. We call this "undesirable disequilibrium." However, if the visitor finds that they have access to resources that can help, they will be more likely to embrace their disequilibrium, continue to engage with the exhibit, and have the opportunity to experience productive struggle. It is the difference between these possible paths that led us to conclude that the Invite stage is essential to the design framework. It is important to note that designing for this stage is not intended to be any different from what exhibit professionals always strive for; in fact, for a long time we called this stage "just good design." Museum professionals have a wealth of institution-specific knowledge about how to design welcoming and accessible exhibits, and we have included some relevant references that our team has found helpful in the resources table at the end of this section.

The framework presents several overarching strategies within each of the approaches. Productive struggle exhibits should attend to **all** of these strategies.

Minimize barriers to entry: Making it easy to get started with an exhibit allows visitors to devote their "struggle budget" (a person's limited physiological resources to engage with disequilibrium) to the designed challenge rather than learning how to use the exhibit.

- **Provide easy orientation:** When a visitor first walks up to the exhibit, make it easy for them to figure out how to use it. This might involve using interfaces that visitors are familiar with from outside of a museum context, like monitors, buttons, and graphics. Titles, use graphics, and screen-based instructions for all aspects of an activity can help provide easy orientation as well. It can also help to consistently use a specific button shape for a specific task (e.g., our team uses small squares for audio on/off). Using light-up buttons to indicate a visitor's selection can be a helpful strategy, too.

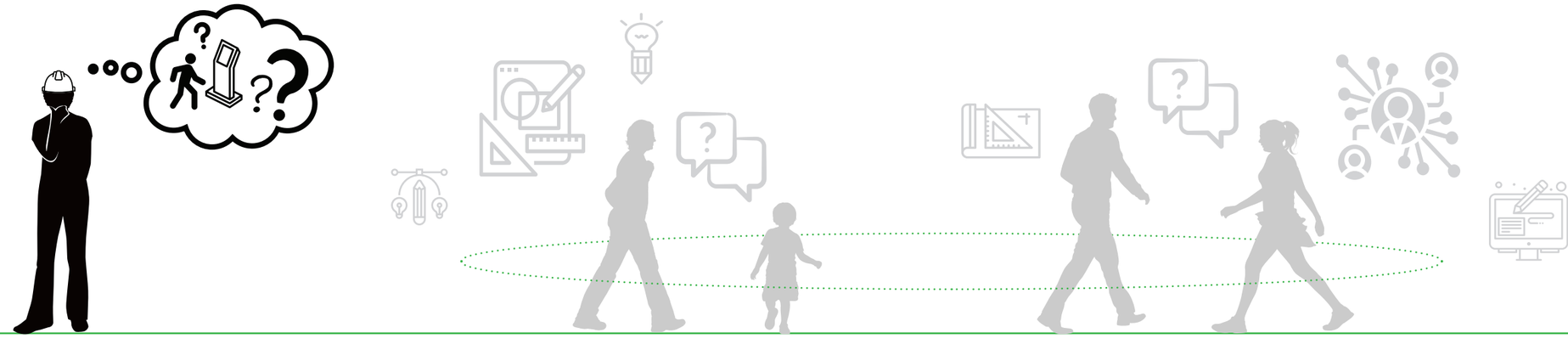
- **Demonstrate clear objectives:** Not only should people know how to use the exhibit, but visitors should be able to figure out quickly what they are trying to do. An exhibit may have multiple possible goals that visitors can pursue, or the exhibit can invite visitors to think up their own goals. However, the key is that at least one main objective is evident soon after arriving at the exhibit. This might be achieved by articulating the objective in the title of the exhibit, using clear language on all labels, or highlighting places or objects that are important to pay attention to.

- **Allow visitors to preview:** This strategy draws from the Exhibit Design for Girls' Engagement (EDGE) framework (Dancstep & Sindorf, 2016), and has been effective for our work on productive struggle. In practice, it means designing an exhibit where one visitor can watch another visitor use the exhibit before they try it themselves. This helps visitors understand what the exhibit is about, how to use it, and whether or not they find it interesting before they decide to engage in a setting where others are able to see their performance. It can also allow visitors to think in advance about what they wish to try when it is their turn. One part of achieving this is to avoid situating the component in a hidden corner or creating a theater-like environment that visitors cannot peek into.

- **Make it obvious how to reset or continue:** Visitors are unpredictable. Prior research at the Museum of Science showed that visitors often think an exhibit is broken if it was designed for everyone to begin their experience at a specific introductory panel or screen, and they do not see that introductory material when they arrive (Kollmann, 2007). Visitors can also be frustrated if pieces are missing from an activity. Make it clear how to start no matter what state the activity is in when a visitor arrives. This may mean including an ever-present start button or ensuring the activity is designed so that a visitor can build on others' prior work.

- **Present a compelling task:** If visitors are not invested in the exhibit, they are unlikely to persist through any disequilibrium they encounter. This strategy applies to making sure there are motivating goals for visitors to pursue as well as ensuring that the actual process of pursuing those goals is something that visitors wish to do. One strategy to achieving this is to story-test the subject matter or activity design with visitors to gauge their interest early on in the development process.

- **Maximize relevance, value, and authenticity:** This design strategy is drawn from CAST's Universal Design for Learning (UDL) guidelines (2018), and is vital for productive struggle. Visitors should be able to connect the activity to their everyday lives, see the activity as relevant to them, develop useful skills, and connect their experience to real-world applications. Formative testing with a diverse group of learners can help exhibit professionals successfully achieve this design strategy for a wide range of visitors.



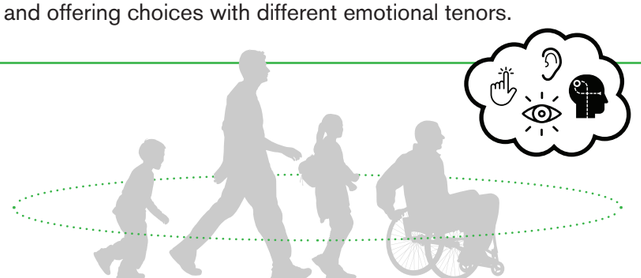
Prioritize accessible design for all: Employing accessibility practices makes productive struggle experiences feel safer and more welcoming for a broader swath of visitors.

- **Design physically inclusive interactions:** Although there are clear guidelines about recommended measurements for reach, pull-under, width of walkways, etc. (CAST, 2018; Majewski, 1996; Museum of Science, 2016), it takes time and effort to make exhibits that meet all these guidelines and ensure the widest possible range of visitors can comfortably use a space. At the time of our project, clear guidelines are not yet available for some exhibit elements (e.g., touch screens), and so require thoughtful experimentation to develop accessible solutions. One strategy is to create multiple versions of an exhibit component (e.g., at different heights) to enable people to use their preferred modes of engagement (e.g., sitting or standing), rather than trying to reach all visitors with a single component. These strategies demand particular attention when a productive struggle exhibit seeks to promote disequilibrium by challenging fine or gross motor skills (see *Disrupt* section).

- **Incorporate multisensory features:** To ensure that different people can maximally engage in productive struggle, think about how visitors can use the activity with multiple senses, without having one sense dominate others. For instance, a non-sighted visitor should not feel like they are missing out on a key element of the activity. Consider how to balance audio, tactile, and visual aspects of learning. When appropriate, having on/off switches for audio, or being able to control the pace, replay, or skip forward through audio and/or video can help visitors tailor their own experience.

- **Avoid reliance on pre-existing skills and specialized knowledge:** To welcome a wide audience, it is vital to make sure everyone can do an activity whether or not they have a particular background or aptitude in the subject. This design strategy is effective for visitors of varying cognitive abilities, schooling backgrounds, ages, and interest in the content area. Important considerations for this design strategy include: using simple vocabulary, defining specialized terms, and including diagrams, graphical representations, and broadcast audio in order to reduce reliance on reading.

- **Provide for varied emotional preferences and skills:** People vary in their emotional preferences and their abilities to manage different emotional states. Offer opportunities for visitors to self-regulate their emotions by choosing different options to fit these needs. Within an activity, this might mean offering chances to take a break or transition to a different task, and offering choices with different emotional tenors.



Disrupt

In this stage of the framework, we focus on three ways to spark feelings of disequilibrium that can spark feelings of disequilibrium: **crafting novelty** (visitors encounter something that seems new or unexpected), **leveraging uncertainty** (visitors are unsure about how to do something or what will happen next), and **introducing social unease** (visitors' experiences go against their culturally constructed expectations).

In this stage of the design framework, the goal is to initiate disequilibrium. For our team, this stage feels most different from our typical exhibit development approaches. We often want to make exhibits as straightforward as possible, but in this framework we purposefully aim to do the opposite: we intentionally create challenges and design ways to encourage visitors to persist in those challenges to achieve satisfying outcomes. In some cases, creating a challenge may mean being strategic in how you offer support. In some instances, you might make supports available only when visitors choose to use them, rather than offering them by default (we discuss this further in the *Support* section). Or, you may make an easy task harder by leaving some supports out. We encourage you to experiment and see what happens! Visitors are often quite resourceful, and relying on them to take on a little extra effort can make an experience more meaningful overall. We recommend having a target age group for productive struggle such that visitors in that age group are at the sweet spot for difficulty. Then, you can design so that all people of other ages can still enjoy the exhibit, rather than striving for just the right level of challenge for everybody.

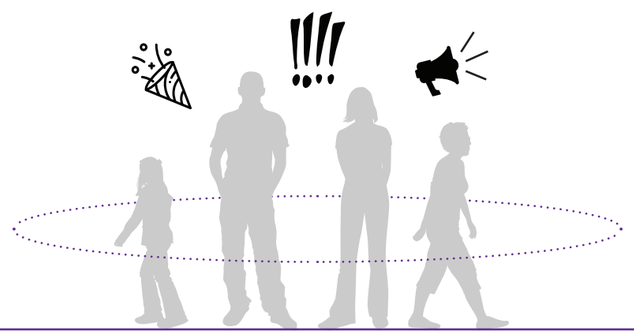
To spark disequilibrium, we recommend applying **at least one** of the three approaches and testing to see whether it is effective. Each approach can be implemented to varying degrees, so testing helps identify when the exhibit is creating enough, or too much, disequilibrium.

Craft novelty: Disequilibrium can arise from encountering unexpected, unfamiliar, or surprising information that challenges visitors' existing concepts.

- **Challenge expectations:** This strategy takes advantage of the fact that people come to a museum with prior experiences and expectations about how things work. For example, in our "Mystery Skulls" exhibit, we found that visitors spontaneously identified an armadillo skull as a dolphin. To support disequilibrium, we included this expected (but incorrect) answer in the options visitors can choose from when identifying the skull. Another approach is to establish expectations in one task and then change them in the next task. For instance, in our "Sneak" exhibit, a certain set of tools is available in the first level of the activity but different tools are available in the next level, forcing visitors to adjust their strategies.

- **Embed surprising phenomena, experiences, or events:** Science museums often have surprising artifacts and interactives in their halls already, which professionals can leverage to support productive struggle. An exhibit can focus visitors' attention on these unintuitive features to highlight disequilibrium. For example, the Coandă Effect can make a ball float in midair in a way that is visually captivating and surprising. Our "Air" exhibit highlighted this phenomenon and encouraged visitors to apply it in order to move a ball over a curved tube into a target.

- **Include unfamiliar information:** Sometimes disequilibrium can arise from learning new things, especially if those things stand at odds with visitors' prior knowledge. For example, in our "Sneak" exhibit, the result of a successful attempt is the uncovering of new information: you thought you were sneaking up on a bird but then discover other animals and learn that other animals listen to birds' alarm calls to learn about approaching danger. If you can move through the forest without startling birds, you are likely to see other animals, as well.

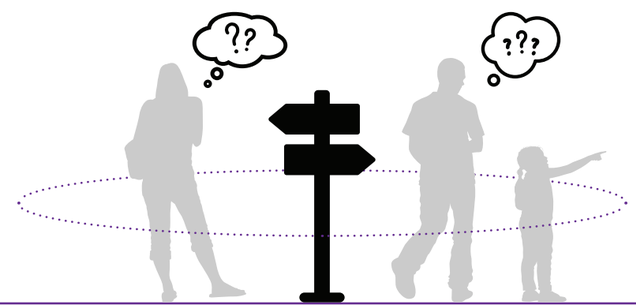


Leverage uncertainty: Visitors can experience disequilibrium when there is ambiguity about what will happen next.

- **Limit available information:** Rather than providing all relevant content to visitors up front, an exhibit can prompt disequilibrium by holding back some information and encouraging visitors to fill in gaps on their own. For example, in "Mystery Skulls" we purposefully omitted information that might have helped visitors answer questions correctly on a first try, and instead incorporated this information through hints that only became available if visitors answered a question incorrectly.

- **Force decision-making:** Requiring people to put a stake in the ground, even if they are not sure of an answer, can be an effective means of sparking disequilibrium. People like to be right, so the simple act of making a choice can increase investment in an activity. For example, after selecting a skull to identify in "Mystery Skulls," visitors are shown photos of three different animals and the exhibit asks them to select which animal they think the skull belongs to before beginning the classification activity. The addition of this one design strategy transformed the emotional tenor of this activity from dull to productive struggle.

- **Challenge fine or gross motor skills:** An exhibit can support disequilibrium by asking visitors to do something that is physically difficult. For example, our "Sneak" exhibit asks visitors to move very slowly to avoid scaring wildlife, and our "Air" exhibit requires careful manipulation of air flow to solve precise challenges. For physical tasks, prototyping with visitors with limited mobility and dexterity is especially important to offering challenges that welcome engagement by the widest possible range of visitors.



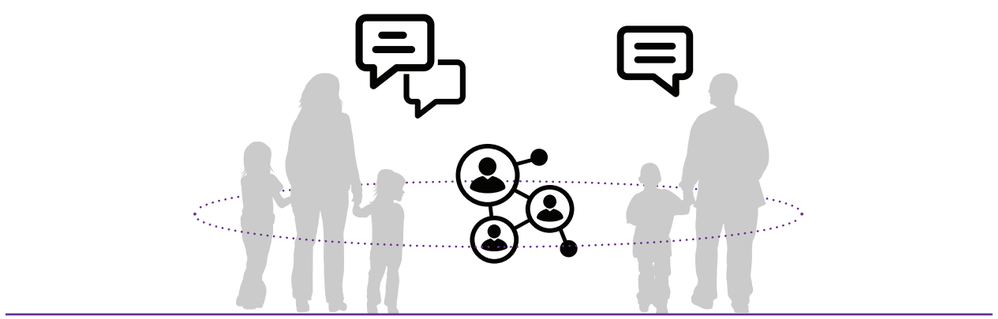
Introduce social unease: Exhibits can foster disequilibrium by challenging cultural norms or encouraging social interactions that make visitors feel uneasy.

- **Invite competition:** Competitive activities can heighten feelings of disequilibrium, so having options to compete with other visitors can be a valuable tool for productive struggle. Competition can occur between group members, with visitors whom you do not know, with a computer, or even with yourself (e.g., topping a previous personal best).

- **Break social norms:** Societies have expectations about how people should interact with one another, whether written or implied. Breaking these expectations can contribute to disequilibrium. Think about how an exhibit might invite a visitor to do something that defies a stereotype or moves beyond typical modes of group engagement in a museum. For example you might take advantage of the fact that most people like to be viewed by others as competent: in an activity where the goal is content learning, guessing incorrectly might be perceived as a judgment of one's intelligence, but, of course, we recognize that sometimes being incorrect is an important part of learning science.

- **Embrace interpersonal differences:** Exhibits can spark disequilibrium by offering multiplayer activities that foster disagreement. In some exhibits, simply having the option of working with another person is likely to bring out interpersonal differences. You might try including a discussion question likely to elicit the sharing of different opinions, offering prompts that encourage people to talk through their varied perspectives, or assigning visitors contrasting positions and inviting them to debate.

- **Offer a performative element:** Sometimes disequilibrium can come from having other people watch us, which can make us self-conscious. In public spaces like museums, you are often in someone's sight line, whether it is your own group members or strangers who happen to be at a museum while you are visiting. Designing an exhibit to emphasize awareness that others are watching can support disequilibrium by making the activity feel like a performance. Our "Sneak" exhibit is a physical performance, while our "Mystery Skulls" exhibit involves a performative demonstration of problem-solving skills.



Support

The framework outlines three approaches for supporting visitors through their disequilibrium: **offering feedback** (so visitors know how they are progressing in an activity), **giving choices** (so visitors can tailor the activity to their needs), and **supporting self-regulation** (to assist visitors in practicing valuable emotional skills).

Our initial framework included separate stages for the persistence and productivity aspects of our productive struggle definition, but our data repeatedly showed that these two aspects of productive struggle were deeply intertwined. Visitors often told us they were motivated by small successes or by the anticipation of success. They felt successful because they persisted through challenges. And often, once someone succeeded, they re-entered a new productive struggle experience, a trend we call mini-arcs (see *What is Productive Struggle?*). As a result, we ultimately combined these two elements, and the intimate, cyclical relationship between persistence and productivity and how these experiences are often cyclical is now represented in the framework as the Support stage.

There is also a critical relationship between Support and Disrupt. Museums are free-choice learning environments; people can leave if they feel like they do not have the resources they need to manage their disequilibrium. Yet, too much support reduces the challenge and the resulting disequilibrium. Further, our research showed that increasing disequilibrium was often associated with greater productivity—the harder the task, the more satisfying it was to complete, assuming the appropriate support was there to make completion possible. We now know that achieving productive struggle requires balancing across design strategies—both within each stage and between the three stages—through prototyping and iteration.

Testing and prototyping are key for finding the right level of support. We recommend that museum professionals do **as many as possible** of the things in the Support section to increase persistence and productivity without diminishing disequilibrium.

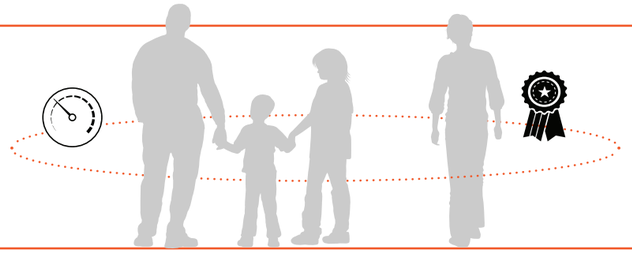
Offer feedback: Exhibits can support persistence and productivity by helping visitors monitor their progress throughout an activity and alerting visitors to the possibility of a preferable outcome when appropriate.

- **Indicate progress or success:** Design can encourage persistence by helping people know where they stand in relation to completing a task, while having a clear start and end helps visitors judge how much farther they have to go. Using clear language to indicate a "good job" or "success" on computer based interactions is a useful strategy that signals productivity. We designed “Mystery Skulls” to incorporate all of these elements. For example, we created a field journal feature that helps visitors keep track of which skull features they have already observed (progress), and we clearly indicate when answers are correct (success).

- **Include mini-wins (throughout) and final rewards:** Make people feel good about themselves by rewarding them for their effort! Exhibits can embed small rewards (which our team calls "mini-wins") throughout an experience as well as more substantial rewards after successful completion of a task. Rewards can involve anything from fun noises and visual flair to learning new facts or getting glimpses of the final reward. People are motivated to work towards rewards (persistence) and people feel good when they earn them (productivity).

- **Integrate hints and scaffolding:** Think about how much information visitors need up front, and how to make additional information available to visitors when they might need it. This ensures that people can find the information they need, but because people often try to avoid using hints, it encourages them to sit with their disequilibrium and attempt to figure things out on their own first—which can be highly satisfying. Careful consideration and testing can help determine what to call hints (e.g., “take a closer look” or “learn more”) in order to ensure that visitors do not feel their intelligence is threatened when they use the hints.

- **Encourage trying again:** Although this might seem like an obvious strategy for supporting persistence, many productive struggle research participants indicated that the reason they stayed at an activity after a failure was because the exhibit instructed them to try again. Encouragement can also come from other group members, so an exhibit can invite supportive social interactions in addition to using labels that invite trying again. For instance, a sign can directly say, “try again” or might more subtly suggest iteration through saying something like “there is a better answer.”



Give choices: Allowing visitors to adjust the activity to best match their own preferences can encourage persistence and lead to feelings of productivity.

- **Offer more than one level of challenge:** What is difficult for one person is not always difficult for someone else, and people will experience disequilibrium at different levels of difficulty, so the more you can provide tasks with varied levels, the more people you are likely to engage in productive struggle. When providing multiple levels of challenge, order matters. Productive struggle is more likely to occur when the difficulty level ramps up over time rather than when easier tasks follow difficult ones. Many visitors start with items that are labelled “easier,” objects that are furthest to the left, or anything tagged with the number "1." Adding layers of difficulty as the visitor moves toward the right, or up from one, can enable productive struggle.

- **Create pathways for social and solo interaction:** Some people are more motivated and rewarded by taking on a task by themselves and feeling fully responsible for its success. For other people, it is preferable to have someone else to collaborate with and share results with. While productive struggle occurs for both individuals and groups, visitors draw on different resources in an exhibit depending on whether they are alone or with others. Formative testing can help identify what resources need to be present for an exhibit to foster productive struggle no matter how many visitors are present.

- **Design for multiple goals:** Having a variety of objectives for an exhibit will help more people find something to do that is motivating and rewarding. Open-ended exhibit design invites visitors to come up with their own tasks, as well as those intended by museum professionals. An exhibit might offer a number of different challenges to try, and could have several ways to succeed at a given challenge. For instance, at one of the challenges in our “Air” exhibit, there are multiple target points where you can try to direct a ball.

- **Allow repeated attempts:** Productive struggle exhibits are intended to be hard, and sometimes things do not go as well as visitors intend, especially the first time. As such, productive struggle exhibits should encourage persistence by designing challenges that visitors can try multiple times. This can involve a visitor repeating a level, or going back to another part of the exhibit to refresh knowledge or skills before returning to a challenge.

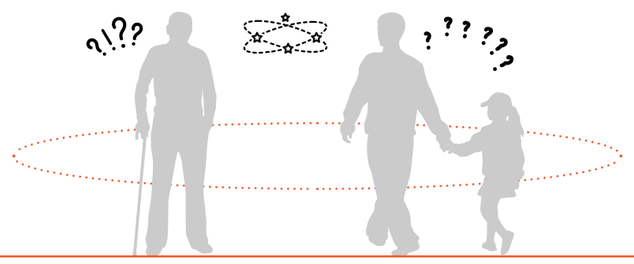
- **Include the option to do less/more:** Different people will find different things satisfying. Productive struggle exhibits should embrace feelings of productivity—when they occur—by offering multiple decision-points about whether to stop or continue. While some people will not feel productive until they do all of the available aspects of an exhibit, others might face significant disequilibrium at just one piece of an exhibit and might wish to stop after resolving their disequilibrium at that component. By allowing visitors to decide how much they want to engage, more people can leave feeling satisfied.

Support self-regulation: Exhibits can be designed to help visitors manage their emotions so they can persist through disequilibrium to a satisfying result. Exhibits might directly acknowledge or name specific emotions so as to normalize feelings like confusion, frustration, and surprise, or indirectly acknowledge them by using images or language about the experience of challenge.

- **Acknowledge the challenge:** Everyone will experience a different level of challenge at a productive struggle exhibit, but it can be valuable to explicitly indicate to visitors that the exhibit can feel difficult. Because many exhibits are designed to be fun and not especially difficult, visitors may hold an expectation of this norm. In experiencing a challenge or difficulty, there is a risk that people could walk away from a productive struggle experience feeling like they are unintelligent or unskilled. Exhibits can acknowledge challenge by labeling a certain level or task as “harder” (implying that it could be difficult), or describing a task as a challenge.

- **Normalize feelings of disequilibrium:** This means making sure visitors know that the experience of disequilibrium (including feelings of confusion, frustration, or surprise) is normal, intentional, and alright! Exhibit signage can normalize disequilibrium with words (e.g., "it’s ok to be frustrated!") or images. In our prototyping, we experimented with signage that featured images of faces that portrayed emotions commonly associated with disequilibrium emotions. While ultimately not included in the final design of our exhibits, the data showed it was effective so we encourage others to try this strategy when appropriate. Normalizing disequilibrium feelings suggests to visitors that other people are likely experience those feelings when doing the exhibit tasks, too.

- **Invite reflection on disequilibrium:** Thinking and talking about your feelings can help you manage them. Exhibits can prompt visitors to reflect on how they felt at an experience and encourage visitors to talk to others about their experiences. For example, an exhibit can include a mechanism for inputting emotional self-report such as selecting among emotions that a visitor might have felt or asking visitors to report their perceived level of difficulty of a task (imagine creating a “challenge-o-meter”!).



Beyond a single exhibit: Considerations for gallery-level design

The design framework focuses primarily on exhibit-level considerations (how to design a single exhibit interactive that elicits productive struggle). However, we know that the broader context that surrounds an individual exhibit can also play a role. Some things you might consider at a gallery-wide level are:

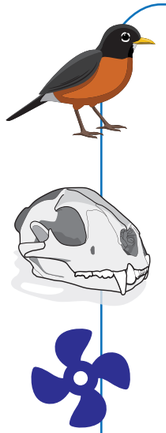
- Supporting easy orientation by using consistent label strategies and interfaces across a gallery;
- Making multiple copies of an exhibit to reduce unwanted frustration due to crowding or premature departures by visitors who want to give others a turn; and/or
- Balancing the energy levels across an exhibition by considering the placement of specific exhibits within the gallery; this can be as simple as making sure there are quiet benches near more active activities.

Resources: Frameworks for designing inviting exhibits

Audience	Abbreviated Citations (see *References* for complete information)
Girls	Dancstep (née Dancu), T. & Sindorf, L. (2016). *Exhibit designs for girl’s engagement: A guide to the EDGE design attributes.*
Families	Borun, M. (1998). *Family learning in museums: The PISEC perspective.* Gaskins, S. (2016). Children’s learning in museums with their families.
Visitors with disabilities	CAST. (2018). *Universal design for learning guidelines.* Majewski, J. (1996). *Smithsonian guidelines for accessible exhibition design.* Museum of Science. (2014). *Universal design plan for exhibit design & development.*
Low-income visitors and ethnic minorities	Dawson, E. (2014). “Not designed for us”: How science museums and science centers socially exclude low-income, minority ethnic groups.
Bilingual visitors	Yalowitz, S., Garibay, C., Renner, N., & Plaza, C. (2013). *Bilingual exhibit research initiative: institutional and intergenerational experiences with bilingual exhibitions.*
LGBTQ+ visitors	American Alliance of Museums. (2016). *Welcoming guidelines for museums.*
All audiences	American Alliance of Museums. (2018). *Facing change: Insights from AAM’s diversity, equity, accessibility, and inclusion working group.* Science Museum Group (2017). *Engaging all audiences with science: Science capital and informal science learning.*

Introduction to Examples from Research & Practice

On paper, a design framework can seem straightforward and easy to implement. In practice, design is a nuanced process that involves identifying and implementing a cohesive suite of strategies that work together to achieve your experience goals. In examining an exhibit as a whole, and considering the ways that specific design features interact to shape the visitor experience, the seemingly clear bullet points in the design framework can quickly become muddy and confusing.



To help others contextualize the framework, the following pages describe how we applied it in three exhibit design processes.

A **Case Example** for each exhibit describes how we came to select the exhibit for inclusion in our research project, key points in the development process for that exhibit, and a narrative discussion of strategies that worked, as well as some of those that we tried but found ineffective for that exhibit.

A **Framework Implementation Example** for each exhibit includes a brief description of the exhibit activity and the visitor experience, and follows the three-stage structure of the design framework to indicate which strategies we employed at that exhibit, along with a description of how each strategy manifested for that exhibit.

As we worked to clarify how the strategies presented by the framework played out in different design circumstances, our own analysis of the three exhibits sparked many conversations among colleagues. The good news for anyone who wants to apply the framework to create a productive struggle exhibit is that it does not matter if a single design attribute contributes to implementing multiple strategies within a stage, or even if some design attributes seem to work to achieve productive struggle across multiple stages. For instance, the same feature of your exhibit might both "challenge expectations" and "include unfamiliar information." The spirit of the approach, which in this case is "craft novelty," can be achieved through a single strategy, or through multiple strategies in combination. Similarly, a design attribute that "invites competition" (a Disrupt stage strategy) might also "allow repeated attempts" (a Support stage strategy). The specific strategies are intimately linked and the impact of individual strategies need not be disambiguated to achieve productive struggle.

For ease of reading, we decided to present each design attribute of an exhibit as it relates to just one of the strategies presented by the framework. In doing so, we mean to define each of the framework strategies using illustrative examples from our work in this project. However, in making this choice, we recognize that the full complexity of how the design attributes worked together to create productive struggle (across approaches within a stage and across stages) is not completely represented. As you review the case studies, keep in mind that these are simplified stories of our process that highlight our decision points. Descriptions of case studies are simplified in two ways.

First, **many of the strategies we used to support the Invite stage are just good design**, reflecting the Universal Design for Learning approach that we utilize in designing any new exhibit. These universally-good-design attributes (shared in the table below) are not explicitly included in our examples, because we expect that a practitioner could apply these in any exhibit context. Our case examples and associated framework implementation examples therefore focus on design attributes specific to the individual exhibits.

Universally-good design strategies for creating inviting exhibits

General

- Keep Universal Design guidelines in mind (e.g., CAST, 2018; Majewski, 1996; Museum of Science, 2014); see www.mos.org/UniversalDesign.
- Create multisensory experiences: tactile, audial, visual, olfactory.
- Communicate messages through multiple modes that support each other.
- Make the goal of the activity obvious at first glance or in a short preview.
- Standardize the user interface and activity flow at each interactive.

Graphics

- Use legible fonts and high-contrast colors.
- Include drawings (use graphics) that visually display how to use the exhibit.
- Avoid red-green color coding, or relying on color alone to communicate information.

Content

- Write clear labels that everyone can understand.
- Avoid technical language and include definitions when necessary.
- Use active voice.
- Use as few words as possible to communicate the content.
- Include a main title label, as well as a sentence or question to introduce the primary message.
- Break ideas into separate paragraphs.

Physical Design

- Ensure pull under, tabletop height, reach, angles of monitors, and positioning of graphic slants are in comfortable positions for someone sitting or standing, someone with low mobility, small humans, and large humans.
- Across a gallery, facilitate orientation by standardizing the use and placement of button boxes, individual buttons, monitors, audio labeling, and other elements that are present at multiple exhibits.
- Use consistent design for features present at multiple components that have the same functionality, such as button shape and size or the approach to audio labeling.
- Standardize iconography of monitors versus touchscreens (on-screen monitor elements should not look like touchscreen buttons).
- Do not include lights that strobe or flicker.
- For full body activities, indicate where to enter, exit, or stand using signage and/or flooring changes (e.g., contrasting carpet area, feet, and arrows).

Second, **designing for productive struggle can be messy, and every exhibit development effort will be different**. In "Sneak," we tried many strategies for Support, but not all of these were included in the final exhibit. In "Mystery Skulls," we tried several Disrupt strategies before finding one that created disequilibrium effectively for the majority of visitors who participated in our testing, and some of the strategies we tried for this stage ended up being effective as Support features instead. In "Air," we expanded the productive struggle experience from a single interactive into a small gallery of interactives that, when experienced together, adhere to the framework. In each case, it was the process of evaluation and iteration that confirmed we had finally achieved productive struggle for our visitors.

Our case examples and framework implementation examples illustrate the iterative design process we used to develop our productive struggle exhibits and how our team defines the design attributes of these exhibits within the context of the framework.

Each development process brought us new insights about how design strategies contribute to emotional experiences. We hope these examples, combined with the framework itself, provide useful guidance for your own design processes as well!

Case Example: "Sneak"

Background

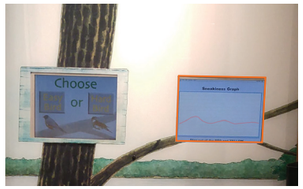
Original "Sneaking Corridor" exhibit (prior to productive struggle redesign)

"Sneak" was the first exhibit we developed for productive struggle. It began as a refurbishment of an existing exhibit, called "Sneaking Corridor," which was up for renewal. Due to the age of the exhibit, the original design did not meet Universal Design guidelines–specifically, it did not have sufficient space for a wheelchair to make the needed 90 degree turn to exit through a set of swinging doors. Also, evaluation had shown that many visitors were not learning key content that we hoped the experience would deliver. Our team saw strong potential for productive struggle in the exhibit. The whole-body nature of the task–moving slowly down a walkway to sneak up on a bird–offered the chance to test ideas about prompting disequilibrium through a physical challenge. Many of the strategies for the Disrupt and Support stages of the framework that we were interested in testing were already present, or could be easily incorporated, within the exhibit design. We anticipated that iterating on this exhibit would allow us to experiment with supports to learn how we might vary the level of challenge in a task (the original exhibit already included two difficulty settings) and how best to offer feedback about progress in an activity (such feedback was also present in the original exhibit, but we saw room for improvement). We were also interested in the potential strategy of "rewards" (in the form of learning something new, experiencing something new, or experiencing recognition from others) to foster feelings of productivity.

We tested the original exhibit with visitors in order to gain a baseline understanding of how it was, or was not, supporting productive struggle. We found some evidence that visitors were experiencing disequilibrium. However, much of the disequilibrium arose for unintended reasons, such as visitors not understanding how to use the exhibit. We also found that visitors were not feeling much satisfaction–the exhibit's content goals were unclear to visitors and many were entirely missing the intended "reward" at the end of their experience.

Problematic Features of the Original Exhibit

A Sneakiness Graph communicated a visitor's speed as they was travelled down the walkway.



Successfully sneaking up on a virtual bird would reveal a short video clip of a deer on a small TV screen at the end of the corridor, located near the floor and just out of the primary user's view.

- Several labels, intended to both guide visitors through the activity and connect the act of sneaking to content about birds' alarm calls, were located in multiple places within the sneaking corridor and on its exterior.

Through eye-tracking data, we found that few visitors even noticed the TV screen. Further, surveys and interviews informed us that the intended content ("go slow to keep the bird from noticing you, sounding an alarm, and scaring the animals away") was not clear, as few visitors read all of the various labels. Finally, instead of observing the virtual bird's vocalizations and body language, we found that visitors primarily paid attention to the Sneakiness Graph; although this captured visitors' attention, we also found that some visitors did not understand what the graph was intended to communicate.

We hoped that by enhancing and clarifying these existing exhibit features we could transform this exhibit to maximize the potential for productive struggle.

Creating Productive Struggle in "Sneak"

Improving the Invitation

First, we improved the physical layout of the exhibit to better welcome to better welcome all visitors and to clarify the goal of the activity.

The updated exhibit, called "Sneak," iterated on the original design to make it more inviting to visitors. These changes addressed the undesirable disequilibrium that we saw in preliminary testing, allowing people to focus on the challenge of sneaking up on the bird.

- The walkway was widened and the swinging gates at the exit were removed to ensure wheelchair access.
- The original version already had good visibility to preview other visitors trying it out before your turn; those sightlines were extended in the new design.

While it was clear to visitors what they needed to do (move cautiously down the walkway), the motivation for that task (reveal a deer at the end of the walkway) was not. Discovery of the deer was also meant to convey key content: you can see more outdoor wildlife if you avoid detection by birds, whose alarm calls alert other animals that you are coming. This design provided visitors with little motivation or reason to anticipate this "reward".

We also re-designed the digital interface to create a sense of anticipation for visitors.

- We moved the deer animation and other key content to a single, larger screen, prominently located at the end of the walkway.
- We also added animated bushes that slowly move apart as you progress down the walkway. First you see just bushes, then a deer's tail, and then finally, at the end (if you are successful) the deer are revealed. If you move too fast, you only see the deer as they run away.

Crafting Supports for Persistence and Productivity

In addition to making the physical layout more welcoming, and consolidating all the feedback onto a single digital screen that is clearly visible at all times, we had also iterated on the design of each feedback mechanism and the "rewards" offered.

For example, the main mechanism through which visitors gained feedback about their progress in the original exhibit was a Sneakiness Graph that showed the visitor's velocity as a function of time. This graph was not easy to read or interpret. We redesigned it with a more familiar form factor (a speedometer), which became the "Sneak-o-meter." The bird's signals to visitors were always multisensory (audio and visual), but we further emphasized the visual and auditory feedback in the new exhibit using enhanced animations and directional speakers.

Persistance and Productivity in "Sneaking Corridor" and "Sneak"

Original Exhibit	Revised Exhibit
Visitors choose between the "easy bird" (robin) or the "hard bird" (woodthrush).	Visitors choose between the "easier bird" (robin) or the "harder bird" (woodthrush).
A Sneakiness Graph displays your velocity as a function of time on one of the two screens at the end of the corridor.	A Sneak-o-meter (like an odometer) displays your current speed on the single screen at the end of the corridor.
On the second screen, bird calls and animated body language indicate whether you are moving too quickly. The birds pause their singing when you are detected. If you move too fast, their alarm call scares the hidden animal away.	On the same screen, bird calls and animated body language indicate whether you are moving too quickly. The birds pause their singing when you are detected. If you move too fast, their alarm call scares the hidden animal away.
The exhibit is set in a forest scene that includes trees and bushes; a TV monitor with the reward is hidden in the bushes at the end of the corridor, near the floor.	The exhibit is set in a forest scene that includes trees and bushes; animated bushes on the screen at the end of the corridor slowly part as you progress, eventually revealing the reward.
Upon success, a reward prompt states: "Look down to your left." and a short video clip of a deer plays on the TV monitor near the floor. The prompt is the same for both birds.	Upon success, the reward prompt for the robin states: "Good sneaking! You discovered a mother deer and her fawn." and displays an animation of two deer interacting. The reward prompt for the woodthrush states: "Good sneaking! You discovered a family of deer." and displays several deer interacting.
Failure is indicated by a flashing, red light on one of two screens at the end of the corridor.	A failure screen states: "You were too fast. The robin's call alerted the animal and it ran away! Exit and try again," and displays a brief glimpse of the hidden deer as they run away.
Levels are based on velocity sensitivity; all supports are present for both levels.	Levels are based on velocity sensitivity; the Sneak-o-meter support is removed at the "harder bird" level.

Both versions of the exhibit had two difficulty settings: an "easier bird" (with a lenient velocity sensitivity setting) and a "harder bird" (with a stricter sensitivity). In the final version of the exhibit, visitors have the support of the Sneak-o-meter only if they select the robin ("easier bird"). For the wood thrush ("harder bird"), visitors must rely on the bird's vocalizations and body language to indicate if they are moving too fast. We also varied the reward for each level: if visitors are able to sneak by the robin, they see a mother deer and her fawn in the clearing; if visitors are able to sneak by the wood thrush, they see an entire family of deer in the clearing.

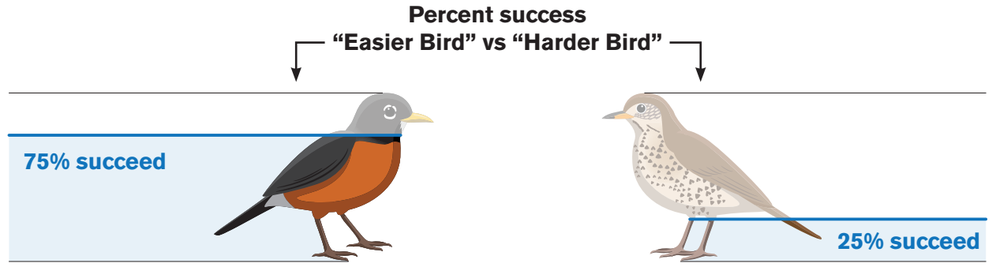
A Lesson Learned

More support does not always make the task easier.

At first we thought that we could affect the exhibit's level of challenge by allowing visitors to choose which types of supports they would like to have: audible bird vocalizations, a Sneak-o-meter, and bird body language (e.g., turning to look at you, calling, or flying away). We even did a round of testing where visitors were given different combinations of these supports with the hypothesis that taking them away would help increase the exhibit's challenge. Participants in this testing did the activity twice—once with all three supports and the other time with a combination of just two supports (we randomly assigned the order). We found that visitors perceived the second version of the activity they experienced as harder, no matter which version they tried first—it was adapting to new conditions that created the difficulty, rather than the presence or absence of any particular support!

"Sneak" taught us that tasks can feel harder if the information available to you changes.

Although we ultimately decided to provide all three supports for the "easy bird" and to remove the Sneak-o-meter for the "harder bird", we knew from our testing that the order in which visitors tried the two birds would matter more for their perceived difficulty than the presence or absence of supports. So, instead of relying on the supports to dictate the level of difficulty and associated disequilibrium, we turned to changing the objective challenge of the activity. Using Kinect technology, we adjusted the exhibit to require visitors to move more slowly at the harder level. We made changes to the sensitivity and tracked the resulting success and failure rates at different settings. Ultimately, we set the Kinect such that (on average) 25% of attempts at the "easy bird" would fail and 75% of attempts at the "hard bird" would fail. At these settings, visitors clearly differentiated the difficulty between the two birds. In final testing of the exhibit, this design approach reliably elicited productive struggle, with close to 90% of participants reporting disequilibrium, persistence, and productivity.



Framework Implementation Example: “**Sneak**”

Visitors learn to practice their sneaking skills and attend to bird cues in order to better observe animals in the wild.

New “Sneak” exhibit
(after productive struggle redesign)

Invite

Use clear design to welcome all intended learners to the activity.

Minimize barriers to entry:

- ***Provide easy orientation:*** High contrast entry/exit signage and instructions that share how to use the exhibit orient visitors to the activity.
- ***Demonstrate clear objectives:*** The exhibit title is “Sneak” (which is the main goal of the activity) and all graphics feature a bird or set of birds.
- ***Allow visitors to preview:*** Visitors can observe other users prior to engagement.
- ***Make it obvious how to reset or continue:*** The activity automatically resets after a successful or a failed attempt.
- ***Present a compelling task:*** Visitors enjoy the challenge of moving their bodies slowly and learning how to pay attention to bird calls and other cues.
- ***Maximize relevance, value, and authenticity:*** The behaviors of the bird and deer in the activity mimic those that are observable in nature; other naturalistic features (e.g., plants and rocks) make the overall experience feel welcoming by simulating the outdoors.

Prioritize accessible design for all:

- ***Design physically inclusive interactions:*** Audio labeling provides orientation to the physical design; the activity is wheelchair accessible; graphics are clearly legible and have high-contrast; and activity instructions and other information is provided in multiple ways (audio and visual).
- ***Incorporate multisensory features:*** Directional speakers play bird calls (audio) and the digital screen (visual) reacts to visitor’s movement (physical).
- ***Avoid reliance on pre-existing skills and knowledge:*** Visitors do not need to have any previous knowledge of birding or observing animals.
- ***Provide for varied emotional preferences and skills:*** A button box allows visitors to choose between an “easier bird” and a “harder bird.”

“Sneak” Visitor Experience:

In the re-designed interactive, visitors enter the “Sneak” walkway and use a button box to select either an “easier bird” or a “harder bird.” At the end of the walkway, a large screen shows an animated forest scene with a bird on a branch. An on-screen prompt encourages visitors to slowly sneak up to the bird to discover animals that are hidden behind the bushes. As the visitor moves slowly towards the screen, the bushes and trees in the animation gradually part to reveal a clearing in the forest. The bird’s calls and behaviors change in response to the visitor’s motion. Bird song indicates non-threatening, slow movements. Silence and a sharp head turn let visitors know they are moving quickly enough to draw the attention of the bird and they must slow down. An alarm call from the bird means the visitor moved too fast and the activity is over. For the “easier bird”, additional feedback is provided by a Sneak-o-meter (which is similar to a speedometer); but the sneak-o-meter is not present for the “harder bird.” Visitors who successfully sneak up to a bird discover an animation of a deer family in a clearing on the screen. If the visitor is not successful in sneaking up to the bird, the bird’s call alerts the hidden animals and visitors only see the animals as they run away. Throughout the experience, other group members can observe your progress, your successes, and failures.

Disrupt

Facilitate disequilibrium by challenging norms or expectations.

Craft novelty

- ***Challenge expectations:*** When going from the Sneak-o-meter (“easier bird”) to lack thereof (“harder bird”)—or vice versa—visitors have to adapt to the change in supports to understand when and how to move.
- ***Embed surprising phenomena, experiences, or events:*** The bird stops singing and abruptly turns to look at the visitor if they make sudden movements, and makes a sharp alarm call when visitors move too quickly.
- ***Include unfamiliar information:*** Visitors encounter evidence that birds make alarm calls when startled, alerting animals in the area that there might be a threat.

Leverage uncertainty

- ***Limit available information:*** The “harder bird” does not include Sneak-o-meter support, so visitors must rely only on the bird’s visual and audio cues to guide their sneaking.
- ***Force decision-making:*** Not featured in this exhibit’s design.
- ***Challenge fine or gross motor skills:*** Visitors must control their full body and pay close attention to their own movements to achieve success.

Introduce social unease

- ***Invite competition:*** Individuals in a group can compete to see who can achieve success the fastest, in the fewest attempts, or in creative ways such as crawling or creeping along the wall.
- ***Break social norms:*** The slow sneaking behavior that visitors must engage in to be successful can feel unnatural.
- ***Embrace interpersonal differences:*** Visitors can choose any number of ways to sneak down the walkway, from tip toeing to crawling.
- ***Offer a performative element:*** Visitors must complete the full-body activity independently, in front of other visitors in the gallery.

Support

Provide options for persisting through disequilibrium and feeling productive.

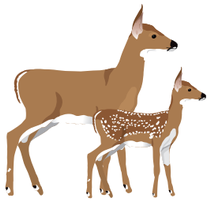
Offer feedback

Indicate progress or success: A visitor's progress is indicated as the bushes on the screen part incrementally. The bird's vocalizations and body language indicate whether visitors are moving at an appropriate speed or too fast. For the "easier bird," the Sneak-o-meter also indicates a visitor's speed of movement (the meter is color-coded with green, yellow and red zones).

Include mini-wins (throughout) and final rewards: Visitors get glimpses of the clearing ahead as the bushes slowly part. If successful, they are rewarded with a final reveal of a deer family in the clearing and the final screen reads "Good sneaking!"

Integrate hints and scaffolding: The Sneak-o-meter and animated bird (it's vocalizations and body language) help visitors learn to slow down.

Encourage trying again: When the activity concludes (after either success or failure), the screen encourages visitors to "Exit and try again."



Give choices

Offer more than one level of challenge: There are two levels of difficulty: a robin ("easier bird") and a wood thrush ("harder bird").

Create pathways for social and solo interaction: Only one visitor can sneak at a time, but other members of their group can observe and interact with the person who is sneaking.

Design for multiple goals: Although the main task is sneaking up on a bird, visitors frequently set their own goals, such as sneaking up on the bird while keeping the Sneak-o-meter in the green zone, completing the activity by crawling rather than walking, or being successful but moving as quickly as possible.

Allow repeated attempts: Visitors can repeat the activity as many times as they choose.

Include the option to do less/more: Visitors have the option to try only one bird or both birds.

Support self-regulation

Acknowledge the challenge: Labeling the different challenges "easier bird" and "harder bird" suggests that a higher level of difficulty is expected.

Normalize disequilibrium: Not featured in this exhibit's design.

Invite reflection on disequilibrium: Not featured in this exhibit's design.

Case Example: "Mystery Skulls"

Background

We developed "Mystery Skulls" to replace an exhibit called "Mammal Skull Mysteries," which encouraged visitors to practice classification skills of observing and grouping. Visitors explored a set of five physical, 3-dimensional skulls using a computer interface, which presented a series of yes or no questions about various skull features and provided information about animals' diets and lifestyles.

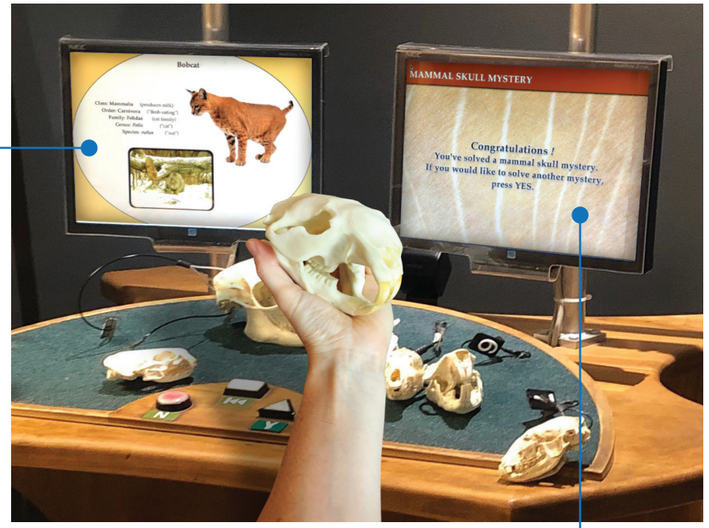
Original "Mammal Skull Mysteries" exhibit (prior to productive struggle redesign)

While keeping the goals of the original exhibit in mind, we wanted to develop a new version that would address the original exhibit's flaws and elicit productive struggle. The team suspected that we could prompt disequilibrium (specifically, surprise or unease) through careful selection of the skulls that would be included in the new version of the exhibit. We also thought we could improve both persistence and productivity by providing more meaningful support for visitors' observations of the skulls and individual skull features, leading to an exhibit that would feel more satisfying overall.

Problematic Features of the Original Exhibit

The computer was not able to track which skull the visitor was observing, so the prompts provided to the visitor were the same regardless of which skull they had selected.

The computer was not able to track any mistakes visitors made when answering the yes or no questions about the skull features, so visitors received no feedback about their answers.



Sometimes content about specific animals was presented to visitors, even though this content did not necessarily relate to the specific skull a visitor was actively observing.

On the final screen, visitors were congratulated for correctly identifying a skull, regardless of how they had answered the questions (and with no connection to which skull the visitor had been observing).

We tested the original exhibit and found that there was little evidence that visitors were experiencing desirable disequilibrium, experiencing a sense of productivity, or were otherwise emotionally engaged in the activity or invested in its outcome. We hoped to re-design the activity flow to more explicitly connect screen-based prompts to the specific skull a visitor was observing. In doing so, we believed we could prompt desirable disequilibrium, support visitors to persist in observing skull features, and increase feelings of productivity for visitors (by confirming, at the end of the activity, whether they had ultimately identified the correct owner of the skull).

Creating productive struggle in “Mystery Skulls”

Improving the Invitation

First, we addressed flaws in the original exhibit that hindered basic usability and could lead to misconceptions about skulls.

- We redesigned the computer interface so that visitors use a button to select the skull they wish to explore and are guided through an investigation of that specific skull.
- For each skull, visitors answer a series of questions about three of that skull’s features, instead of a series of generic questions about skulls.
- The computer keeps track of the visitor’s answers, and offers a hint if the visitor initially answers a question incorrectly.
- Visitors are presented with a screen that reviews all of their (correct) answers before making a second, final guess about what animal the skull belongs to.

These changes dramatically improved the overall user experience compared to the original exhibit. Visitors could track their own decisions and progress in the activity, rest assured the computer’s prompts were relevant to the skull they were observing, and would no longer be congratulated for misidentifying a skull and walk away with misconceptions. However, prototyping of this version of the activity did not reveal strong evidence of productive struggle; the activity was emotionally bland.

Crafting Supports for Disequilibrium and Productivity

A key aspect of the redesign was creating a computer interface that offers support based on the specific skull a visitor choses to explore. The new version of the exhibit invites visitors to first make an initial guess about what animal a skull belongs to, presenting them with a choice of three possible animals. Visitors then answer questions as they observe the skull’s features and access hints that help them learn how each feature is related to how an animal lives. Finally, visitors piece together this information–along with their prior knowledge–to confirm or disconfirm their initial hunches when they make a final guess, before the true identity of the skull’s owner is revealed.

Would unusual looking skulls elicit disequilibrium?

To get to this final activity flow, we spent significant time getting the challenge “just right” for the majority of visitors. In a series of prototypes, we tried sparking disequilibrium in a variety of ways, including by:

- presenting unusual skulls (with unique characteristics or from uncommon animals);
- presenting skulls that “break the rules” visitors learn as they explore other skulls; or
- forcing an initial guess, prior to visitors’ scaffolded exploration of skull features.

In each attempt, we encountered something unexpected that challenged our ideas: re-designing this exhibit was a productive struggle experience for our entire team!

A Lesson Learned

Sparking disequilibrium can also increase feelings of productivity.

Early in our team's conversations we hypothesized that including skulls with dramatic or unusual features would elicit disequilibrium. We believed that observing skulls, visually and tactilely, would be a novel experience for many visitors, and that by including skulls that looked unfamiliar or even "creepy" we could further heighten an emotional sense of surprise or unease. After choosing a variety of potential skulls with unique characteristics (e.g. unusual teeth, unexpected textures or colors, extreme size, or unique protrusions such as horns), we conducted preliminary prototyping to determine visitors' reactions. Our testing revealed that, rather than promoting disequilibrium, including such skulls made our activity too easy for visitors. The skulls we considered "surprising" were in fact most easily identified, likely because their unusual features were often reflected in distinctive elements of the living animal's appearance, making it easy for visitors to guess correctly.

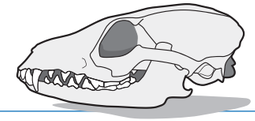
In our next iteration, we tried to encourage disequilibrium by increasing the perceived difficulty of the activity. The team had previously discussed that disequilibrium might be sparked by opportunities to learn new or unexpected information. To this end, as visitors investigate the skulls and observe individual features, they learn "rules" for each feature (e.g. eyes on the sides indicate a prey species); but, to prompt disequilibrium, we purposefully included skulls that present exceptions to these rules (e.g. the Gila monster has eyes on the sides of its head, but it is classified in this activity as a predator).

We knew from testing at "Sneak" that increasing the perceived difficulty of the challenge in subsequent attempts can support productive struggle, and we had learned during early "Mystery Skulls" prototyping that visitors typically elected to explore skulls from left to right. We decided to design for increased difficulty by purposefully placing skulls that broke the rules further to the right, increasing the likelihood that visitors would learn a rule before encountering the exception. While observing and classifying skull features might have gotten easier for subsequent skulls without implementation of this design strategy, we found that considering the natural tendencies of visitors' free-choice explorations could help support productive struggle. Despite these changes, as testing continued we gathered evidence that the activity did not seem to create productive struggle—fewer than one-third of participants experienced disequilibrium and productivity.

The team convened to unpack these results: making incorrect guesses did not seem to cause disequilibrium and making correct guesses did not seem to prompt feelings of productivity. In fact, receiving feedback about their guesses did not seem to matter much to visitors at all! This suggested that visitors' interest in learning new information was not sufficient motivation for deeper engagement, so we discussed options for boosting visitors' emotional investment in the activity. Ultimately, we redesigned the flow of the activity to require visitors to make a guess about which animal the skull belonged to at the beginning of the activity. This approach challenged our instincts as exhibit developers; welcoming and inclusive design typically avoids test-like experiences in which visitors may not know the answers. However, we anticipated that forcing a guess could raise the stakes for visitors by piquing their natural curiosity to discover whether their guess was right or wrong, motivating them to attend to information about the skull features as evidence that could confirm or disconfirm their hunches. In selecting animals to present as trios of possible skull owners, we ensured the choices were believable by including skulls that were similar in shape and placement of features. Early testing had revealed some of visitors' hunches about what animal each skull belonged to; for example, many visitors spontaneously guessed the armadillo skull came from a baby dolphin, so we included a dolphin among the trio of possibilities for this skull.

"Mystery Skulls" taught us that forcing an initial guess and offering content that "breaks the rules" can spark disequilibrium; these strategies also seemed to increase visitors' investment in the ultimate outcome of the activity, leading to increased feelings of productivity when the final answer was revealed.

In the next round of testing, we found that this change–forcing visitors to make an investment in the activity by putting forth (and recording!) an initial best-guess–was critical to creating productive struggle. We found that all participants experienced disequilibrium in our testing of the final version of the activity. We also saw that, when disequilibrium increased, more visitors (100% in this final testing) experienced productivity as well. We posit that the heightened investment in the activity made it feel more worthwhile overall.



Persistance and Productivity in "Mammal Skull Mysteries" and "Mystery Skulls"

Original Exhibit	Revised Exhibit
Visitors must keep track of the skull they have chosen to observe.	A button press tells the computer which skull the visitor has chosen to explore.
Visitors begin a free-form exploration of skull features without any investment in the outcome of their investigation.	The activity flow requires visitors to make an initial guess (before a scaffolded exploration of that skull's features) and a final guess (once an investigation of skull features is complete), raising the stakes for visitors.
The interface presents a series of generic yes/no questions about skull features, providing no feedback on visitors' answers.	The interface presents a series of skull-specific questions about features, offering a hint, and an opportunity to try again, when visitors answer a question incorrectly.
Visitors must keep track of the skull features they have observed and what they learn from those observations.	The computer tracks the features that have been observed, and visitors' (correct) answers about those features, keeping a running summary of those observations. The interface presents a summary of skull feature observations for review, before prompting the final guess.
At the end of the question set, the interface always congratulates visitors for successfully identifying a skull, leaving some visitors with misconceptions due to a lack of feedback.	If visitors select an incorrect animal in their final guess, they have the opportunity to make additional guesses. A congratulations screen appears when the correct animal is selected, along with additional information about the animal.

Framework Implementation Example: "**Mystery Skulls"**

Visitors identify animal skulls by observing skull features, learning relationships between these features and how an animal lives, and noticing patterns or groupings across animals.

New "Mystery Skulls" exhibit
(after productive struggle redesign)

Invite

Use clear design to welcome all intended learners to the activity.

Minimize barriers to entry:

- ***Provide easy orientation:*** Screen-based prompts guide the visitor through the activity.
- ***Demonstrate clear objectives:*** The exhibit title and numbered skulls orient visitors to their overall task.
- ***Allow visitors to preview:*** Visitors can observe other users prior to engagement.
- ***Make it obvious how to reset or continue:*** A button box clearly communicates button functions (e.g. start over, audio on/off, enter, and left/right selection buttons).
- ***Present a compelling task:*** Visitors report that they enjoy observing and identifying animal skulls.
- ***Maximize relevance, value, and authenticity:*** The activity uses casts of real animal skulls and teaches skills that scientists use.

Prioritize accessible design for all:

- ***Design physically inclusive interactions:*** Audio orientation to the physical design is provided (via audiophone); the exhibit is designed to comply with ADA standards of reach and wheelchair pull-under; and all skulls are on turntables allowing 360° rotation and easy exploration for low dexterity visitors.
- ***Incorporate multisensory features:*** Text and graphics are designed with high contrast, legible font; the exhibit features broadcast audio for screen readout, with an on/off toggle; the skull number buttons are raised to improve accessibility for blind and low-vision visitors; and skulls are cast with mouths open to maximize available tactile information.
- ***Avoid reliance on pre-existing skills and knowledge:*** The activity is designed such that all the necessary information is embedded in the exhibit.
- ***Provide for varied emotional preferences and skills:*** Multiple skulls, features, and supports allow visitors to make choices to customize their experience.

"Mystery Skulls" Visitor Experience:

In the re-designed interactive, after selecting one of the five skulls to explore, visitors are immediately prompted to make a guess about which animal the skull came from, choosing from among a trio of possible animals. After their choice has been recorded, visitors use the digital interface to explore up to three features of the skull. As they explore, visitors closely observe each skull feature to learn more about how the animal lived and are provided with feedback and clues that help differentiate the skull from the other two animals in the trio. After visitors have explored one or more features, they can choose to "Solve the Mystery" and use the information they learned about each of the features to make a final guess about which animal in the trio they think the skull belonged to. Throughout the experience, additional hints are offered whenever visitors choose an incorrect answer, and visitors are encouraged to try again.

Disrupt

Facilitate disequilibrium by challenging norms or expectations.

Craft novelty

- ***Challenge expectations:*** After visitors have learned guidelines about skull classification, the activity challenges these expectations with skulls that "break the rules" (i.e. the Gila monster has eyes on the side, but is classified as a predator).
- ***Embed surprising phenomena, experiences, or events:*** The activity includes animal skulls that have distinguishing characteristics and may be perceived as unusual.
- ***Include unfamiliar information:*** Visitors learn facts and terminology for skull classification (i.e. the sagittal crest is a ridge of bone along the top of the skull that can indicate the strength of an animal's bite).

Leverage uncertainty

- ***Limit available information:*** Some information is available only as hints, which can be accessed only with an additional click after an incorrect guess, encouraging visitors to try to puzzle through the questions with limited information before relying on scaffolding.
- ***Force decision-making:*** After selecting which skull to identify, the visitor is shown three pictures of animals and must guess which animal the skull belongs to before beginning the activity.
- ***Challenge fine or gross motor skills:*** Not featured in this exhibit's design.

Introduce social unease

- ***Invite competition:*** Not featured in this exhibit's design.
- ***Break social norms:*** The tendency to want to avoid failure heightens the stakes of this activity, as there are numerous questions that visitors can get right or wrong.
- ***Embrace interpersonal differences:*** When working in a group, visitors may have differences of opinions about which answers are best, and they have to persevere and work through that together.
- ***Offer a performative element:*** Visitors are required to answer questions in front of other visitors.

Support

Provide options for persisting through disequilibrium and feeling productive.

Offer feedback

Indicate progress or success: After answering a question correctly there is a "great job" screen and conclusions from each feature are displayed. At the end, when a visitor selects the correct animal, there is a success screen.

Include mini-wins (throughout) and final rewards: Visitors have the opportunity to feel successful as they answer questions about features, revise their guess about the animal, or identify individual skulls.

Integrate hints and scaffolding: If the first question about a feature is answered incorrectly, visitors have the option to "take a closer look" where a hint suggests which part of the skull to look at. If the second question about a feature is answered incorrectly, visitors can choose to "learn more" about the specified feature.

Encourage trying again: Throughout the activity, visitors are presented with a "try again" button when they select incorrect answers. Visitors who make an incorrect final guess learn that "there is a better answer"—text on this screen acknowledges similarities between the animal selected and the mystery animal, but also describes what differentiates them.

Give choices

Offer more than one level of challenge: There are five different skulls to choose from, including skulls that "break the rules". The difficulty of the skulls increases progressively from left to right (the order in which our testing indicated most visitors explore the skulls); because rule-breakers come later in the experience, questions about any given skull feature also become progressively harder as visitors explore additional skulls.

Create pathways for social and solo interaction: The activity can be accomplished alone or in a group.

Design for multiple goals: Visitors can set goals based on their interests, such as completing all of the skulls with unusual features, or repeating a skull until they answer all of the feature questions correctly on the first try.

Allow repeated attempts: Visitors can repeat a question, a feature or a skull, that they have already completed.

Include the option to do less/more: After exploring at least one feature, visitors can choose to identify the skull at any time. Visitors can do as many or few skulls as they choose.

Support self-regulation

Acknowledge the challenge: The activity acknowledges when rules are broken and provides an explanation.

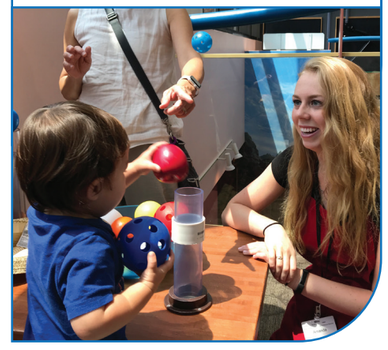
Normalize disequilibrium: Not featured in this exhibit's design.

Invite reflection on disequilibrium: Not featured in this exhibit's design.

Case Example: "Air"

Background

For our third productive struggle exhibit, we embarked on building an exhibit from scratch. The team was committed to studying both natural history and physical science exhibits; since "Sneak" and "Mystery Skulls" are both natural history experiences, we wanted to work on a physics exhibit. We were fascinated by what we saw when visitors engaged with phenomena of air, because it seemed characteristic of disequilibrium. In our *Discovery Center*, young children and their families regularly expressed surprise at our "Air Table" component, as they tried to get objects of different shapes, sizes, and weights to float in a stream of air and manipulated that airflow using pipe attachments of varying diameters. In our *Science in the Park* exhibition, we often saw visitors persistently puzzling through how to throw a beach ball into a stream of air so it would float suspended within the airflow. However, our content experts lamented on inaccurate labels (across many science centers) that claim a ball hovers above a blowing fan because of the Bernouli Effect, when these interactives actually demonstrate the Coandă Effect! The popularity of our own version of this exhibit prevented its removal, but there had not yet been an opportunity to make updates that would effectively share the correct explanation for this phenomenon with visitors. Both of these air-related exhibits were scheduled to be replaced by upcoming renovations, but we knew they were visitor favorites and prime candidates for a productive struggle makeover. We suspected there was something about the emotional tenor associated with the phenomena that we could not only replicate, but amplify. Therefore, our third productive struggle exhibit would focus on the surprising (and sometimes confusing) properties of air.



Original air exhibits

Problematic Features of Previous Air Exhibits

As we began our planning process, we were confident that we could use air to spark disequilibrium, but there were two key obstacles that we anticipated from the start. First, our past air exhibits were traditionally open-ended and exploratory. We saw some visitors make their own goals, but sometimes it seemed like people walked away with more of an "oh, that was neat" feeling, rather than a sense of "yes, I accomplished something!" We suspected that presenting a surprising phenomena on its own would not be enough to support productive struggle for all visitors. How could we support those feelings of productivity? Second, when we thought about the Invite stage, we were concerned about employing traditional interpretation strategies for air (e.g., graphically visualizing the presence/movement of air), but excited about the potential to provide multisensory opportunities to observe air flow, particularly for visitors who are blind or have low vision. How could we make an exhibit about air fully accessible?

Creating Productive Struggle in “Air”

Our process began with brainstorming all sorts of things you can do with air, with a focus on generating specific challenges that would feel satisfying to visitors who completed them. The list of potential challenges quickly grew, and we decided that we would prototype a variety of challenges to see which ones were most effective. Having many different tasks, however, provided a new challenge for us: how would we design an experience in which the different tasks hold together as a coherent productive struggle experience?

“Air Basics” introduces visitors to the phenomena that can be used to solve a series of challenges.

Crafting Supports for Disequilibrium and Productivity

Our team quickly decided that all of the exhibit’s activities would involve using air to move ping pong balls to different targets. To accomplish the activities, visitors would need to know something about air, but we recognized that giving the solutions away too easily could detract from productive struggle.

Drawing on the framework strategy “limit available information,” we designed “Air Basics” as a separate, library-like reference area where visitors can explore the various phenomena (produced with combinations of fans) that you would need to solve a set of “Air Challenges.” “Air Challenges” would each require visitors to apply more than one of the phenomena represented by “Air Basics” activities in order to move a ball to a desired target. Visitors could start with “Air Basics,” learn the foundational principles, and then go apply them to the “Air Challenges” activities. Alternatively, visitors could start with the challenges and use the “Air Basics” section only when they got stuck and needed help.

New “Air” exhibit (after productive struggle redesign)

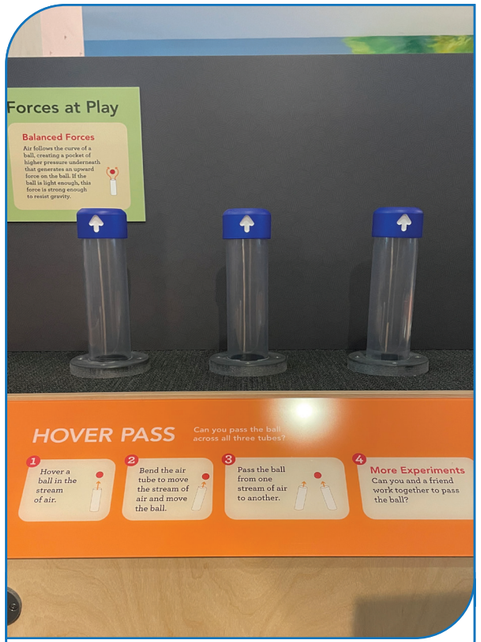
Our research confirmed our hypothesis about this approach: having tools available on-demand (but not automatically provided) gave visitors agency to modulate their own “just-right” level of difficulty.

Balancing Disequilibrium and Productivity - “Hover Pass”

As we iterated on the “Air Challenges” activities, we worked to carefully balance the need to rely on prior skills or knowledge to solve the challenges with our goal of creating experiences that could be both welcoming and satisfying for any visitor. We wanted the activities to be difficult enough to spark disequilibrium, but not so difficult that people would walk away. Complicating our design effort, airflows in the exhibit were sometimes hard to control because the blowers and vacuums that we built were always interacting with other external forces. This meant that, even if you understood what you were supposed to do, the ball did not always behave exactly as you expected because of potential interactions with the HVAC airstreams, humidity levels, or other visitors’ locations in the gallery. At some level, we liked this variability, because it seemed too easy if the “challenge” was just another iteration of an “Air Basics” activity. On the other hand, one of our early prototypes was so hard and unpredictable that only our technical designer, who had spent hours making and testing the challenge, was able to complete it–and even his success rate was low! We knew we needed to find a middle ground.

We found that our “Hover Pass” challenge was particularly successful at offering both solo and social interaction opportunities. At this activity, three flexible, vertical tubes are embedded in a horizontal table. Each tube has a fan under it. The goal is to place a ball in the airstream above one tube and bend the tubes in order to pass the ball from one airstream to another. It is possible to do this activity on your own, but it is designed such that you can accomplish the challenge quicker if you work with another person.

As we iterated to find this balance, we also dug into the social elements of our design framework. We gathered evidence about the strategy “create pathways for social and solo interaction.” While all of our final “Air Challenges” offer both solo and social interactions, we found that “Hover Pass” was particularly successful as both a social and a solo experience.



“Hover Pass” challenge

During our testing, we were satisfied to see that “Air” successfully supported productive struggle for many visitors. However, we also found evidence of unproductive disequilibrium that revolved around usability challenges and not understanding the direction of the air flow within the components. This required us to revisit the Invite stage of our framework.

Improving the Invitation - "Direct"

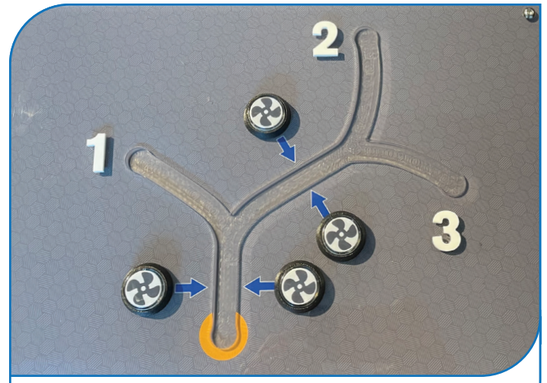
When we looked across our suite of potential activities and the data from visitors who tried them, some activities were successful and needed only minor adjustments. Others were so problematic that we decided to scrap them entirely. We also saw that, as we'd suspected, the invisibility of air and complex interactions of airflows (within and outside of the exhibit components) made the "Air" exhibit (undesirably) challenging from an accessibility perspective. Luckily this project provided an opportunity for us to explore new approaches to universal design for exhibits that we had never been able to pursue in the past.

Our "Direct" activity, specifically, encouraged us to think beyond our traditional scope. In this activity, visitors turn on select fans to dictate the direction of airflow in a pinball-like table maze. When a ping pong ball is dropped into this maze, it follows the airstream and, ultimately, is directed into a target by the airflow. Typically, we rely on tactile representations for visitors who are blind or low-vision, but in "Direct," the fans and paths operate under plexiglass and the balls move very quickly. We placed tactile reliefs on the top of the plexiglass so visitors could orient themselves to the activity by feeling the edges of the paths and targets, thereby learning where the fans were and which directions they pointed. Yet, there was no real-time feedback about where the ball was as it moved through the maze, other than visual observation of the ball itself.

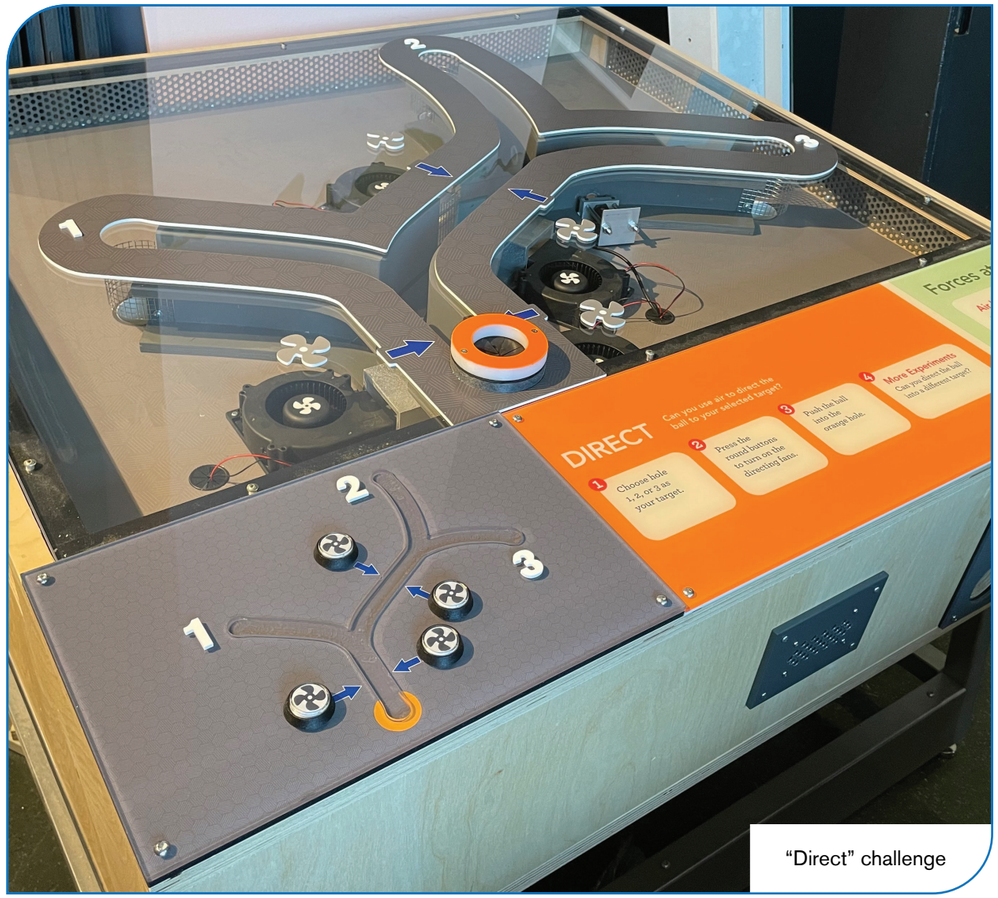
We therefore developed a concept for an audio-tactile map, a small-scale version of the maze that highlights key features of the interactive (the location of the fans, the entry to the maze, the paths balls can take, and exit points from the maze) using both tactile and auditory cues. This map helps orient visitors to the activity: as visitors draw their fingers along a path, they hear the tones associated with that path. Then, when a ball is placed in the maze, a series of beam breaks trigger the same tones as the ball moves along that path, allowing the visitor to hear which path the airflow carries the ball down and which target it reaches. We were able to refine this approach through testing with blind and low-vision visitors, and we have found it to be a valuable approach that has enhanced our understanding of how to make accessible exhibits.

With these improvements to the design, our testing showed that the "Air" exhibit was highly effective at supporting productive struggle. In our final study, participants (n=33) reported that they experienced disequilibrium and productivity at the exhibit, with only three of those people indicating that their disequilibrium was primarily associated with figuring out how to orient to the activity.

In the "Direct" challenge, an audio-tactile map supports visitors to track the movement of the ball using tactile and auditory cues. Visitors: 1) trace their finger along the audio-tactile map; 2) select fans to turn on, 3) place a ball in the maze, and 4) listen for tones in the maze that match those traced by their finger on the map.



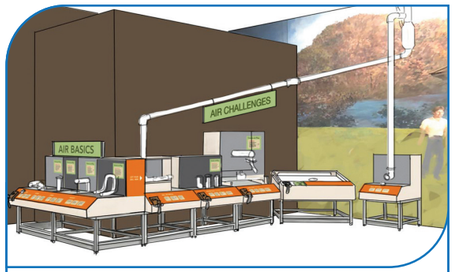
Audio-tactile map



"Direct" challenge

Framework Implementation Example: **“Air”**

Visitors explore the phenomenon of air and the different ways it can be used to move ping pong balls.



New “Air” exhibit
(after productive struggle redesign)

Invite

Use clear design to welcome all intended learners to the activity.

Minimize barriers to entry:

Provide easy orientation: “Air Basics” is on the left of the overall exhibit layout and has a title signaling that it is the place to start. Activities to the right get progressively more complex. Use graphics (pictorial descriptions of how to use the activity) and audio labels introduce each activity; these are positioned in the same location for each one so they are easy to find.

Demonstrate clear objectives: The title signs of “Air Challenges” and “Air Basics” tell visitors what the activity is about. Use graphics provide instructions with clear steps to follow. Blue arrows indicate the direction of airflow, and orange rings show where to place your ping pong ball to begin each activity.

Allow visitors to preview: Visitors can observe one another prior to engagement.

Make it obvious how to reset or continue: Vacuums and fans are either always on or turn on with a button. Vacuums and fans switch off with a timer or by toggle.

Present a compelling task: People love playing with the effects of air!

Maximize relevance, value, and authenticity: Ping pong balls are familiar objects, and everyday vacuums and fans produce the same phenomena that are presented by the components.

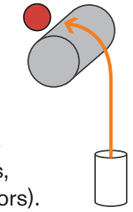
Prioritize accessible design for all:

Design physically inclusive interactions: An audio label orients visitors to the physical design. The activity has side pull-up and lower table tops to accommodate easy reach. Graphics are highly legible with high-contrast color choices.

Incorporate multisensory features: Visitors can feel the direction of air flow into or from the tubes and observe the movement of ping pong balls visually and often auditorily. The challenge called “Direct” has a full-scale, optically clear, tactile map of the activity’s maze on top of the maze, and a scaled audio-tactile map on the control board. Both the full-scale activity and audio-tactile map trigger sounds to indicate which path the ball follows and where it ends up.

Avoid reliance on pre-existing skills and knowledge: “Air Basics” teaches the physics behind the phenomena and lets visitors practice skills they can use on other, more difficult activities (the challenges).

Provide for varied emotional preferences and skills: Visitors can choose from activities that invite different energy levels, including activities that require patience, as well as active tasks like picking up loose ping pong balls and returning them via pneumatic tube to a ball storage tank.



“Air” Visitor Experience:

Visitors encounter a suite of exhibit components that allow them to explore relationships between air flow and object movement. The components are broken into two categories, “Air Basics” (designed primarily as supports) and “Air Challenges” (designed as disruptors). At “Air Basics,” visitors investigate four different ways that air can be used to move a ping pong ball: blowing air over the top of a cup to get a ball to pop out; hovering a ball above a tube of blowing air; traveling a ball around a cylindrical object as it hovers in blowing air; and getting a ball to travel through a tube using suction. Each of these four activities isolates a method for using air to move a ball, revealing the routes that air takes to act on objects in its path and graphically emphasizing these routes. These components are each accompanied by an instruction label, a label that describes the science behind the target phenomenon, and a diagram showing the flow of the air that results in that phenomenon. In a series of three “Air Challenges” components–called “Hover Pass,” “Curve Over,” and “Direct”–visitors must apply the basic techniques that are presented in “Air Basics” to move a ping pong ball from one spot to a target, navigating through activity layouts that introduce obstacles or other additional variables. With multiple activities present (offering both scaffolds and challenges), visitors can choose to explore as many or as few as they like, and the layout of the components allow both social and solo interactions.

Disrupt

Facilitate disequilibrium by challenging norms or expectations.

Craft novelty

Challenge expectations: In many cases, the solution to the challenge defies most visitor’s initial guesses (e.g. the best way to move a ball from one side of a tube to the other is not always to blow air straight onto it).

Embed surprising phenomena, experiences, or events: Activities are centered around how you can manipulate air to move a ball in different ways, including surprising phenomena like the Coandă effect.

Include unfamiliar information: Not featured in this exhibit’s design.

Leverage uncertainty

Limit available information: The “Air Basics” components are physically separated from the challenge components. Visitors who go straight to the “Air Challenges” will confront the limited information available there and may wish to go to “Air Basics” to learn key skills.

Force decision-making: Not featured in this exhibit’s design.

Challenge fine or gross motor skills: Success in many of the interactives relies on the user’s manipulation and the speed and steadiness with which they move. Even for someone who understands the underlying concept, it may take a few tries to succeed.

Introduce social unease

Invite competition: Visitors can compete against themselves or other individuals in a group to see how fast they are able to complete the challenge(s).

Break social norms: Not featured in this exhibit’s design.

Embrace interpersonal differences: Not featured in this exhibit’s design.

Offer a performative element: All visitor actions at this exhibit are on display.

Support

Provide options for persisting through disequilibrium and feeling productive.

Offer feedback

Indicate progress or success: Success is measured by whether you can achieve the desired outcome by using the airflow to maneuver the ball to the intended target. Progress is indicated in the "Direct" challenge through sound.

Include mini-wins (throughout) and final rewards: Each individual challenge has multiple steps, so completing just one step can feel like a mini-win. Succeeding (overall) at an individual challenge is rewarding, and completing all of the challenges might be considered the ulitmate reward!

Integrate hints and scaffolding: Scaffolding is included in the form of "Air Basics." The visitor can return to those activities for additional support at any time.

Encourage trying again: Not featured in this exhibit's design.

Give choices

Offer more than one level of challenge: "Air Basics" activities are easier than the "Air Challenges" activities; the "Air Challenges" also vary in difficulty.

Create pathways for social and solo interaction: All activities can be done alone, and some of the activities can be completed as a team, (e.g., the "Hover Pass" challenge).

Design for multiple goals: The labels include prompts and "more experiments" to try. Visitors are also free to establish their own goals.

Allow repeated attempts: There is no limit to the number of times visitors can repeat an activity.

Include the option to do less/more: Visitors can choose to do as many or as few of the activities as desired, in any order.

Support self-regulation

Acknowledge the challenge: The labeling of "Air Basics" indicates the easier level, implying the harder level of the "Air Challenges" activities.

Normalize disequilibrium: Not featured in this exhibit's design.

Invite reflection on disequilibrium: Not featured in this exhibit's design.

Part III: Measuring Productive Struggle

How Do We Measure Productive Struggle?

How do you know when someone is experiencing productive struggle at an exhibit, and how do you know when your design strategies are helping achieve that goal? This section describes the various methods we used to answer these questions. Over the course of this project, our team utilized a variety of data collection methods, refined them, and built knowledge about how to best leverage them. This involved figuring out ways to collect data about whether or not productive struggle was happening–including evidence of disequilibrium, persistence, and productivity–along with traditional formative data to improve each exhibit's accessibility, usability, and ability to meet its learning goals. While we share a variety of potential methods in this section, one important thing to keep in mind is that investigating productive struggle and applying the design framework does not have to be a complicated process. You can try things out with minimal time, expertise, and materials. For example, our team often does both surveys and interviews as a general practice, but you may find that a survey will suffice for your needs. One size does not fit all, and we invite you to modify our approaches or try new ways to investigate productive struggle.

Throughout this section, we provide examples of questions we asked research participants during our project, describing how we developed and used these questions in our studies. To illustrate how we used various methods in combination, the end of this section presents two examples of data collection protocols used during our project and the methods we incorporated in them. In addition, an *Instrument Appendix* describes protocols, items, and materials our team used when conducting data collection, which you can adapt for your own needs.

Assessing the Outcomes of Productive Struggle for Visitors

This Guide focuses on designing exhibits for productive struggle, and how to incorporate data into the design process. Our project also conducted a final research study about the exhibits we developed, and we encourage you to read our journal articles about the findings. At a high level, we were successful in designing for the emotional outcomes of productive struggle and visitors found productive struggle exhibits to be valuable, educational, and memorable experiences in which they felt like they were doing science. As our articles and other resources become available, they will be shared at: www.informalscience.org/developing-guidelines-designing-challenging-and-rewarding-interactive-science-exhibits

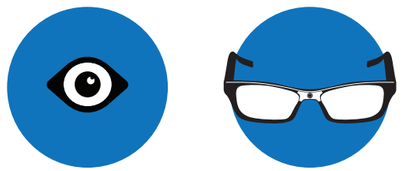
Before you begin, consider how the nature of emotions can affect your data collection.

As described in *Emotions 101*, emotions are complex and are affected by a multitude of internal and external factors and contexts. Not only do the ways we express and articulate emotions vary by individual, but the way we perceive them differs as well–both consciously and unconsciously. Our assumptions about what someone else appears to be feeling may not match that person's internal state. Thus, it is important to incorporate visitor self-report into any approach to measuring productive struggle. One data collection method does not tell the whole story and people vary in their ability to articulate their emotions. As a result, **whenever possible, we encourage using multiple data sources to investigate the extent to which someone is experiencing productive struggle**.

Over the course of this work, we learned that the timing of each aspect of our data collection affected the type of responses we would get from visitors. Memories are quick to change or fade even minutes after an experience, so it is important to ask about emotional experience as soon after an event as possible. Visitors also report different results if you ask them about their overall emotions during the course of an entire experience compared to a specific moment. We found both of these options to be useful, and have developed different approaches for gathering data about in-the-moment emotional events versus when we want to learn about the overall emotional experience of an exhibit. For example, some of our survey questions ask about a visitor's judgment of an entire exhibit experience, while some of our interview methods ask visitors for a moment-by-moment appraisal of their thoughts and behaviors.

This part of the Guide is organized into five sections:

Surveys | Interviews | Guided Recall Activities



Observations | Technology-based data collection

Surveys

Surveys provide an opportunity for visitors to self-report their experiences and perceptions. The surveys we implemented included questions about experiencing emotions related to the productive struggle arc, what aspects of the design supported that experience, and questions from the exhibit team about specific design strategies and outcomes for the individual exhibits. To capture visitors' memories as accurately as possible, our team made sure to gather surveys immediately after visitors used the exhibits. While we included some standard questions across most of our instruments, each instrument was customized to meet the team's needs at the time. Examples of survey questions are grouped below, by question type, and accompanied by a brief description of why we included each type of question.

Assessing an Exhibit's Overall Emotional Tenor

In order to achieve productive struggle, visitors need to feel both challenged and supported by their experience while finding it worthwhile. As emotional experiences cannot be clearly observed, our team found that it made sense to ask visitors how they would categorize their experience overall and we created this survey question:

Which statement best represents how you felt at this exhibit?

- ❑ It felt **easy**, and it was **boring** to keep trying.
- ❑ It felt **easy**, but it was **satisfying** to keep trying.
- ❑ It felt **challenging**, but it was **satisfying** to keep trying.
- ❑ It felt **challenging**, and it was **frustrating** to keep trying.

If you are looking for **an overall sense of whether people might be experiencing productive struggle**, this easy-to-administer question gives a quick gauge of visitors' judgment of their emotional experience at an exhibit. To determine if we achieved our goal, we look for the majority of responses to be "It felt challenging, but it was satisfying to keep trying."

Identifying Productive Struggle Within a Larger Experience

A visitor may have an arc of productive struggle at an exhibit even if the overall tenor of the experience is not perceived by the visitor as "challenging but satisfying." Drawing on a range of affective science literature and open-ended responses from visitors, we developed a list with a mix of productive struggle and non-productive struggle emotions and asked visitors which ones they felt at any point in the activity.

Did you feel any of the following emotions when you were using this exhibit? Circle any emotions you felt:

Confused	Confident	Satisfied	Neutral
Frustrated	Focused	Proud	Relaxed
Challenged	Motivated	Happy	Comfortable
Bored	Pessimistic	Disappointed	Tired

This question can be used to provide **a general overview of the breadth of emotions visitors are experiencing** at an exhibit. During analysis, we looked for the presence of productive struggle words (confused, frustrated, or challenged for disequilibrium; confident, focused, or motivated for persistence; and satisfied, proud, or happy for productivity). We also created space for visitors to record any emotions they felt that were not included among the choices.

Teasing Out Undesirable Disequilibrium

Feelings of disequilibrium may arise for unwanted reasons, such as accessibility or usability problems. This set of questions explores the Invite stage of the framework, to see whether the design strategies for this stage are working as intended.

How much do you agree or disagree with these statements?

	Strongly disagree	Disagree	Agree	Strongly agree
It was easy to figure out **what to do** at this exhibit.	❑	❑	❑	❑
It was easy to figure out **how to use** the exhibit.	❑	❑	❑	❑
The exhibit seemed **broken**.	❑	❑	❑	❑

You can use these questions as a quick way to **determine if visitors are reporting disequilibrium for the right reasons**. If visitors strongly disagree or disagree that it was easy to figure out what to do or how to use the exhibit, or if they agree or strongly agree that the exhibit seemed broken, additional changes may be needed to satisfy the Invite stage of the design framework.

Linking Design and Emotion

In our design-based research process, we needed a quick way to figure out what parts of the design framework were salient to visitors when it was not possible to conduct an experimental comparison between two versions of an exhibit. This led us to create the survey question below, which asks visitors to indicate which design features of an exhibit led to productive struggle feelings. This is intended to follow the example question ("Did you feel any of the following emotions while you were at this exhibit?") so that the data collector only asks about emotions that the visitor felt. The question below lists strategies from "Sneak" that might be related to productivity (Support stage of the framework). For each exhibit we tested, we would customize the list by only asking about design features that were present, and adjusting the language so it was descriptive enough that visitors would know what each item referred to. The *Instrument Appendix* shares a full version of this question, with options targeted for each phase of productive struggle (disequilibrium, persistence, and productivity).

You said you felt proud, satisfied, and/or happy during the activity. What made you feel that way?

- Learning new information
- Seeing the deer at the end (Sneak)
- Having other people at the activity
- Doing the activity myself
- The look and feel of the activity
- Learning a new skill

Our team used this method as a **quick way to map out which design strategies were supporting or not supporting emotional outcomes**.

Interviews

Although they may not always be necessary for your investigations of productive struggle, interviews are a standard part of our team's data collection practice in any exhibit development process. For this project, we used interviews as an opportunity to ask about visitors' productive struggle experiences, perceptions of how the activity's design contributed to their emotions, and exhibit-specific formative questions related to learning and usability.

We almost always used interviews in conjunction with a survey, using the interview questions to gather explanations for visitors' survey responses. This method was **particularly valuable for contextualizing visitors' reported emotional experiences**, related to 1) the overall judgment question, which asks whether the activity was easy and boring, easy but satisfying, challenging but satisfying, or challenging and frustrating; and 2) the questions about whether it was easy to figure out how to use the exhibit, what to do at the exhibit, or whether the exhibit seemed broken. For example, we might remind a visitor that they said the "exhibit was challenging but satisfying" and ask them to explain why.

Could you explain why you feel that way?

Why did you decide to leave the exhibit when you did?

What, if anything, did you learn from this exhibit?

What about this exhibit was confusing or hard to use?

You can also ask the survey questions described in the previous section as interview questions. Early on, we started with a lot of interview questions. Later in the project, we adapted them into survey questions as we began to see common answers that we could use as multiple choice options. This change made data quicker to gather and analyze.

Guided Recall Activities

Our team wanted to learn more about visitors' moment-to-moment experience of productive struggle. This is especially important because our definition of productive struggle depends on the *order* in which emotions occur. To have productive struggle, one must encounter disequilibrium *prior* to persistence, which must happen *before* productivity.

We used two methods to tease apart the order of emotions: storyboarding and stimulated recall. In both methods, researchers ask questions prompting visitors to recall an experience and describe their actions and emotions throughout that experience. The primary difference between these approaches is that storyboarding can be done with minimal tools and takes less time, whereas stimulated recall requires the use of a video recording device and, depending on the length of the exhibit experience, can be time intensive. At their core, both of these guided recall methods are ways to **learn about a visitor's emotions from moment-to-moment**, whereas surveys and interviews are a way to understand a visitor's *overall* emotional experience.

Storyboarding

Similar to writing a comic, storyboarding is a process where researchers guide visitors through a short, paper-based activity to describe a recent experience. Visitors create a step-by-step recollection of what they did on paper cards with short guiding phrases about what they did first, next, and at the end of their experience. We encouraged visitors to use as many "next" cards as they needed to describe their exhibit experience.

First I…	Next I...	Finally I...
...and I felt…	...and I felt…	...and I felt…

After outlining these actions, visitors describe their emotions at each point in their experience by filling out the "I felt…" sections at the bottom of each card. Visitors can write in their own emotions or choose to use a sheet of stickers that present a variety of emotion words–both related and unrelated to productive struggle. As the visitor creates the storyboard, the data collector can probe for more information with questions.

What part of the exhibit made you feel that way? Why? Can you tell me more?

During your experience, I noticed you were using…. What were you thinking or feeling then?

I heard you say… Can you tell me more about what you were feeling when you said that?

Stimulated Recall

Stimulated recall is similar to storyboarding but leverages technology to play back a video recording of visitors' experiences and assist them to recall their experiences during the exhibit activity. As visitors watch the video, researchers prompt them to narrate their experiences. If the visitor is quiet or when there are notable events, the researcher can probe with questions.

What is happening here?

What were you thinking about?

What were you feeling?

What about the exhibit do you think made you feel that way?

In addition to video cameras, our team incorporated biometric data collection using eye-trackers and skin-conductance wristbands alongside our stimulated recall. Researchers used data visualizations of time-series skin conductance peaks and eye-tracking paths to determine when to ask probing questions about visitors' productive struggle experiences (see *Technology-based data collection*).

Observations

Observations are a standard and instrumental part of our exhibit testing practice that help us **learn about visitor behaviors and group interactions**. They can also help us determine specific points in visitors' experiences to probe on during an interview, or identify specific behavior patterns that the design team can reflect on to inform the next steps for exhibit refinement. In general, we conducted our observations with a member of our research team using traditional pen and paper to record notes about visitor conversations and behavior. For some digital exhibit components, we were also able to gather metrics and analytics by incorporating exhibit self-tracking (see *Technology-based data collection*).

Although our observation protocols varied between exhibits, the box below outlines behaviors that our team frequently observed, along with brief notes about how observation of these behaviors helped us assess productive struggle or inform exhibit re-design.

The Utility of Observations in Assessing Productive Struggle

Emotional Experience

Although observation alone is not enough to identify what someone is feeling, some behaviors can be valuable evidence of whether or not productive struggle happens, especially when paired with interviews in which visitors explain what they were feeling. We used observations to consider the implications of:

- **Undesirable disequilibrium**: Do visitors seem to struggle to figure out what to do or how to use the exhibit? Did visitors look at the exhibit and choose not to try it, or leave after being there only a few seconds? This could indicate that visitors are experiencing disequilibrium around orienting to the activity; you might want to revisit the Invite stage.
- **Dwell time and repeated attempts**: How long do visitors stay at the exhibit? Do they try something multiple times, or try multiple aspects of an exhibit? These behaviors could be evidence of persistence; you might wish to follow-up with questions relevant to the Support stage.
- **Conversations**: Do visitors use any emotion words when they talk? Do they ask for help (a potential sign of disequilibrium) or show/tell others what they did or learned (a possible indication of productivity)?
- **Facial expressions and gestures**: Although we cannot tell what someone is feeling when we observe these visible signals, we can ask them about it later, with probes such as, "I noticed you shrugged your shoulders right before you left the activity. What were you feeling when you did that?"

Use of Design Features

We also observed the ways people interacted with the exhibits, which helped us understand what design choices were supporting or hindering productive struggle. We focused on three areas of consideration:

- **Framework strategies**: Look for evidence that connects a design strategy you are using to specific visitor behaviors that are relevant to the stage of the framework that strategy is intended to achieve. Do people read your instructions or use your orientation tools as anticipated (Invite)? Do they undertake the challenge you designed (Disrupt)? Do they draw on the hints or other scaffolding you provided (Support)?
- **Exhibit functionality**: Look for evidence that the design is working as you intended. Does the exhibit work the way you intended? Are there bugs or glitches in any software? Are any pieces broken or missing?
- **Accessibility**: Look for evidence that all visitors can successfully use the exhibit. Do any visitors have difficulty reaching, reading, manipulating, or otherwise using the exhibit materials?

Our team embraces the fact that observation alone is not enough to determine a visitor's emotional experience. How we express and regulate our emotions is both contextual and strongly influenced by social norms, which are affected by our culture. For example, we might assume that a visitor is unhappy if they frown and furrow their brow at an exhibit. However, if we asked them about it later, we may find that they actually felt really focused or surprised instead. Rather than making emotional assumptions based on observations, we noted physical and verbal signs that seemed emotionally-relevant, and used these observations to generate interview probes that asked visitors to reflect on what they were feeling when they displayed these observable behaviors.

Think Alouds

Think alouds are used simultaneously with observations, but encompass aspects of interviews and guided recall activities. This method can provide **a simple way to gather in-the-moment feedback** from visitors while they engage with an exhibit. In this method, visitors are asked to narrate out loud what they are doing, thinking, and feeling as they use an exhibit. The researcher can ask similar probes to those listed in the guided recall activities and this can generate valuable real-time insight. However, it is important to acknowledge that a think aloud interrupts the natural ways visitors use exhibits. It can also act as an intervention itself, as it encourages reflection on the exhibit task and (as the framework illustrates) reflection can support visitors to navigate disequilibrium. Our team used this method while developing an on-line version of the "Mystery Skulls" activity.

"As you use the exhibit, talk outloud about what you are doing. Are there things you find confusing or have trouble using?"

What are you trying to do right now?

I noticed you did X, can you tell me more?

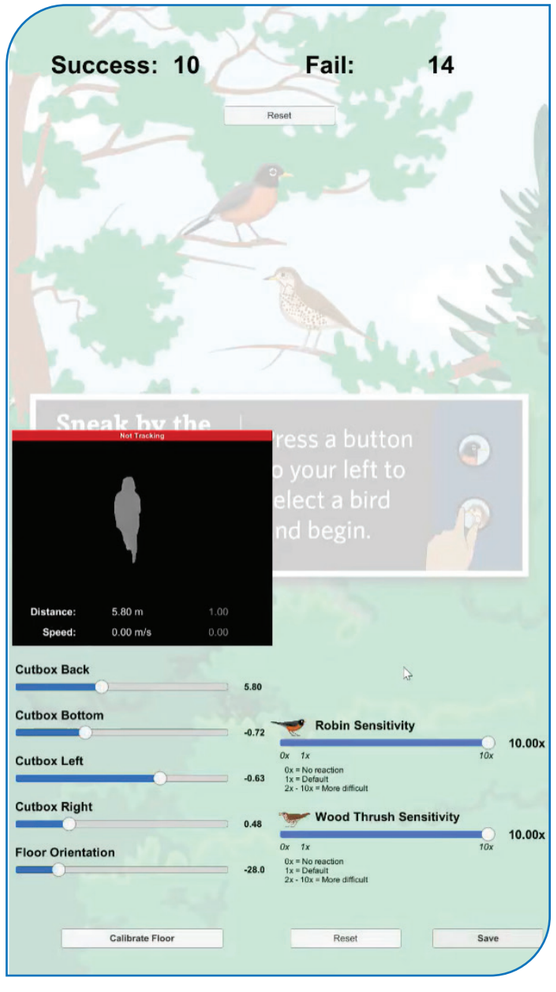
What are you currently thinking and feeling?

Technology-based data collection

Most questions related to whether or not someone is experiencing productive struggle, and if exhibit design is working as intended, can be answered without the use of special equipment. However, biometric technologies such as eye-tracking glasses and skin conductance sensors can present a fuller picture of a visitor's experience. Incorporating embedded analytics and tracking in exhibits can also help reduce the burden of data collection by automating some tasks. Some of these technologies are costly and time-intensive to use. While we used these technologies to answer research questions about the nature and outcomes of productive struggle, they are also valuable for iterating on exhibit design.

Embedded Tracking

Two of our exhibits, "Sneak" and "Mystery Skulls," included digital interfaces, and our in-house interactive media team was able to program data tracking into the software that runs these exhibits. With these tools, we are able to **gather basic data about visitors' interactions with the exhibit–even when no researcher is present.** For example, in "Sneak" we were able to log how many successful and failed attempts people made in trying to sneak up on a bird. We wanted our two levels to be noticeably different, with a high success rate on the "easier bird" and a much lower success rate on the "harder bird." Using the tracking software, we made iterative adjustments to the difficulty levels, tracked data for several hours on each setting, and used the data to ultimately determine the optimal difficulty levels for the two birds. Similar software implemented in "Mystery Skulls" allowed us to track visitors' trajectory through the content paths of the exhibit, which we could compare alongside biometric data and interviews to see if emotional experiences aligned with certain events such as getting questions right or wrong.



Screenshot of the embedded tracking tools for "Sneak." Logs of successes and failed attempts are at the top of the screen, while sliders to change the exhibit's difficulty levels by adjusting the bird sensitivity are in the bottom right.

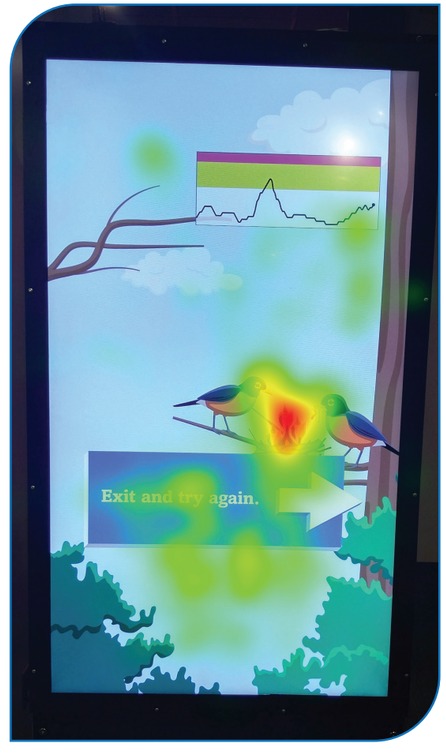
Eye-tracking

Eye-tracking technologies can **gather information on visitors' attention** during an exhibit experience. While screen-based eye-tracking options exist, we opted to use *Tobii* eye-tracking glasses to capture visitors' exhibit experience from their own perspective as they moved around a space.

> *iMotions* software can track the path and duration of a visitor's gaze to create metrics including:
>
> - average time to first fixation (how long it takes for someone to look at something);
> - average gaze time (how long they look at something); and
> - visual heat maps (where they look the most).
>
> Using this data, our team was able to assess what aspects of the exhibit experience visitors attended to and for how long.

Eye-tracking data can provide information about visitors' levels of engagement or disengagement with activities (Rappolt-Schlichtmann et al., 2017). This was particularly useful with early versions of "Sneak" when eye-tracking was able to show the team which on-screen supports visitors were using at the activity. Similarly, early versions of "Mystery Skulls" had two screens but gaze statistics revealed that visitors were only looking at one of them for the majority of the time, which resulted in streamlining the final interface to use a single screen.

While our team found eye-tracking to be valuable, it is time and labor intensive. This includes the necessary work of data cleaning, checking, and analysis setup. Additionally, processing the metrics listed above can take several hours, depending on available computer power. One major lesson-learned was that analysis of gaze data can be complex and time consuming even with specialized software. For formative testing, it may be sufficient to watch the video playback without such analysis.



Heatmap of an early version of the "Sneak" exhibit taken from *iMotions*. The spots in red are areas that visitors attend to the most, while the green is where they looked the least.

Electrodermal Activity

Over the course of our daily lives, our bodies secrete minute quantities of sweat, which change depending on how calm or active we are feeling. This can be measured using wristband sensors that sense our skin's electrical conductivity. Moments of rapid increase in electrodermal activity, known as "peaks," can indicate when visitors feel a change in energy level. Specialized software, such as *iMotions*, uses algorithms to identify peaks based on quantitative criteria. Individuals' patterns of electrodermal activity vary widely, but these algorithms can be helpful in identifying key moments in an experience based on changes from an individual's skin conductance baseline.

Understanding how these changes relate to emotions requires a form of self-report, such as stimulated recall. Researchers can ask probing interview questions to learn what visitors felt in moments with unusual electrodermal activity, such as when there are multiple peaks in succession (Hedman, 2014). Self-report also helps connect emotional experience to exhibit design by allowing visitors to articulate the reason behind their feelings. We may find out that the peaks are directly related to the exhibit design, or find out there was a non-related cause such as a loud sound or excess hand movement creating "noise" in the data.

While electrodermal data can indicate changes in a visitor's feelings, it does not illuminate *what* someone is feeling.

Overall, while they can be challenging to work with, electrodermal activity data provide **a valuable glimpse into continuous, and sometimes unconscious, aspects of affective experience**. Gathering this data requires specialized equipment and analysis software, and the results demand supplemental data collection for meaningful interpretation. However, the information that electrodermal activity provides may not be easily gained through other methods–even the most intensive self-report protocols cannot provide the same level of continuous granularity as a wristband sensor that collects samples many times every second! Furthermore, as described in the *Emotions 101* section, people have varied abilities to articulate their emotions. Electrodermal activity may provide you with insight about experiences that visitors are unable to put into words on thier own.

Image of electrodermal activity taken from *iMotions*. The green shows fluctuations in electrodermal activity, while the pink triangles show "peaks" or moments of rapid increase.

Expression Analysis

We also acquired an emerging technology, *Affectiva*, which analyzes visitors' facial movements as they use an exhibit. The *Affectiva* software uses a computer vision algorithm to detect facial micro expressions (the minute muscular movements that compose a smile, frown, brow furrow, etc.) in video data. This software then makes interpretive predictions about whether facial expressions indicate a positive or negative valence, and analyzes combinations of micro expressions to assess the extent to which the facial movements are consistent with patterns that are commonly associated with specific emotion states like happiness, sadness, or surprise. However, these predictions may not be indicative of a person's internal state (e.g., one can be smiling if uncomfortable or frowning when thoughtful). As with other tech-enabled methods, self-report is still a necessary component in understanding what visitors are feeling and expression analysis should not be used as a sole indicator of a visitor's emotional state.

Image of the "Contemptuous Fern." The dots on the plant represent the technology's approximation of the expression points on a human face, and the orange graph below represents the type of expression the program believes it has detected.

While this technology uses algorithms that offer predictive analysis for several basic emotions, it is still early in development and does not yet offer predictive analysis about more complex emotion states. This includes many of the emotion states our research found to be associated with productive struggle (e.g., confusion, frustration, persistence, or pride). At the time of our project, the technology is prone to both false positives and false negatives. In one test, it identified a plant in one of our exhibits as a face that was expressing contempt! Currently, this technology performs poorly in detecting faces outside of a small range of conditions. Areas of difficulty include exhibits with low lighting or when there is a lot of visitor movement in the area–both of which are common in a museum setting. In addition, ensuring accuracy requires that a person not wear glasses, not have facial hair, and not have a dark skin tone.

For the research described in this Guide, our team was unwilling to rely on a technology that, with its current capabilities, performs poorly in detecting diverse faces–including those of people of color–and was generally unreliable in the museum environment. We ultimately opted not to use this method as a primary data collection tool in our productive struggle studies.

As with electrodermal activity monitoring, our team can imagine the future potential of expression detection technologies to measure continuous data about visitors' emotional experiences. In time, we hope these technologies will improve as companies incorporate more diverse faces into their algorithms' training sets, learn to detect faces at suboptimal camera angles, and more generally improve data capture in dynamic environments such as museums. In tandem, the ethical use of such technologies in public spaces like museums will require critical societal examination of issues of equity, consent, and privacy.

Putting it all together

Now that you have an overview of methods for studying productive struggle exhibits, what does this look like in practice? The table below outlines two protocols our team used. The first is a formative protocol that can be implemented quickly and easily without the use of any technology. This protocol offers a quick check-in about how visitors are using an exhibit's designed features and whether they are experiencing productive struggle; it is meant to be used in iterative cycles of testing and exhibit refinement. The second protocol is crafted as a summative study that focuses on gathering conclusive evidence about an exhibit's ability to elicit productive struggle and identifying the exhibit features that contribute. To provide a more complete picture of visitors' emotional experiences, it includes technology-enabled data collection.

Formative Study; Low-Tech Protocol	Summative Study; High-Tech Protocol
	Pre-survey: Visitors report the emotions they anticipate feeling at the exhibit.
Observation: Researchers watch how visitors interact with exhibit features and note any emotionally-suggestive conversations or behaviors.	**Observation**: Visitors wear eye-tracking glasses and skin sensors; researchers observe behaviors and conversations; the exhibit software tracks analytics; a video camera films visitors' use of the exhibit.
Survey: Visitors report what emotions they felt, their overall judgment of the exhibit's difficulty and satisfaction levels, and (if they felt productive struggle-related emotions) which design features contributed.	**Post-survey**: Visitors indicate what emotions they felt, their overall judgment of the exhibit's difficulty and satisfaction levels, ratings of the exhibit's overall value, and (if they felt productive struggle emotions) what design features contributed to their feelings.
	Stimulated recall: Visitors watch a video of their time in the exhibit and narrate how they were feeling over time; the researcher probes about notable points from the observation and electrodermal activity data.
Interview: Researchers ask for explanations of survey responses and observed behaviors, additional questions related to how the design features contributed to the visitors' experiences, and general questions about learning and usability.	**Interview**: The researcher asks visitors to explain their post-survey responses.
	Final survey: Visitors fill out a survey with demographics and repeated questions from the pre-survey, to see if their responses have changed after using the exhibit.

There are a wide array of methods to choose from when measuring productive struggle. The methods we have described are by no means exhaustive, and the tools listed above are just some examples of what our team used for this project. There is no one-size-fits-all approach and we encourage you to select and adapt the measures that make the most sense for your own work when exploring productive struggle.

What if it's just me? How do I make the most of limited resources?

Although we incorporated a wide range of measures in our studies, we found that you can learn a lot with very little data. This is encouraging because it means that anyone can gather data to inform a productive struggle prototyping process—you need not have research expertise or even copious time to make data-informed decisions to improve your exhibits. In some preliminary testing, you might learn enough to make exhibit improvements by simply asking our close-ended question about overall emotional tenor (whether the activity was easy or hard, and whether it was boring, satisfying, or frustrating) followed by an open-ended, conversational exploration of why visitors felt that way. The most important aspects to keep in mind are: 1) utilizing multiple approaches to capture different aspects of a visitor's experience and 2) the necessity of self-report for understanding emotions.

If you are interested in a more in-depth approach to data collection, but do not feel comfortable tackling it on your own, you might consider partnering with other researchers who have relevant expertise. This may be particularly useful if you are interested in some of the tech-enabled biometric methods described in this section, or if you want to investigate research questions related to emotion and affect but need help navigating because these areas are new to you.

Conclusion

Your heart drops. "What went wrong?" you wonder. "My team and I spent so much time developing this exhibit, and now as I watch visitors use it, I see them getting wrong answers, second-guessing themselves, and feeling confused! I need to re-think my approach. Maybe I should add more hints up front? Or make the task a bit easier?" You pause and take stock of what you're feeling. "I guess I tend to get thrown off a bit when I see visitors having a hard time. I wonder if other exhibit designers ever feel the same way."

Moments of disequilibrium do not only happen when visitors engage with exhibits designed for productive struggle. We can all feel thrown off, uncomfortable, confused, and frustrated at many points in our daily lives and for many different reasons. Members of the productive struggle project team have been no different. Throughout our time on this project, we have felt unsure, surprised, frustrated, anxious, and have experienced many feelings associated with struggle. Research team members have faced technical difficulties related to using high-tech equipment, resulting in lost data. Content developers have felt unsure about so explicitly and purposefully confusing visitors. Designers have puzzled through complex problems related to creating exhibits that make both positive and negative emotions accessible to broad audiences. Our advisors had to manage feeling out of the loop sometimes, as the team's structure, composition and timeline shifted to accommodate organizational change. And across the team, we have all ridden waves of uncertainty about building exhibits based on a framework that was still under development.

Why did we persist? While there are many answers to this question, we would like to highlight just a few, inspired by our design framework.

We found the work compelling. Puzzling through new ideas, making sense of data, creating new modes of engagement, and finding new ways to tap into emotional design, all helped the project team stay the course.

We believed it would be worthwhile. Team members saw value in productive struggle for visitors' engagement and learning, and were also driven by our own interests to advance our practice around emotional accessibility.

We felt encouraged by our progress along the way. Giving ourselves the opportunity to apply The Productive Struggle Framework to three exhibits, each with their own iterative development, played a huge role in helping us see our progressive successes. By the time we developed the "Air" exhibit from scratch, thinking through productive struggle design features felt almost automatic.

We (eventually, and with help) came to accept disequilibrium as a normal part of the process. This work was hard. Fortunately, we had team members well versed in supporting learners (like us!) to recognize, welcome, and navigate these feelings, toward hopefully productive ends.

Were we productive? Were we satisfied with how this work turned out? The answer is yes! And no. In this final section, we share concluding remarks about lessons learned, problems solved, and what we see as the most satisfying results of this process. But we also want to leave you with some of our lingering questions, unresolved problems, and ultimately our suggestions for future directions this work might take. We know there is still more to do. We see the project described in this Guide as a first step toward a future when emotion, in all its complexity, lies at the heart of exhibit design, and we hope these final reflections will inspire new initiatives at your own institutions. In this spirit, we might never see this work resolved completely.

Connecting Theory, Research, and Practice

Throughout this Guide, we have explored how to make sense of and intentionally design for emotion in a number of ways. Theories of emotion helped ground our design and research approaches in evidence-based principles. Our own research and data supported our team to make meaning of the ways design and context played a role in visitors' experiences. And our final design framework was driven by the need for practical guidance for fostering and supporting certain emotional paths. This Guide is written in a way that might suggest that theory informs research, and research informs practice, in a fairly linear way. However, our experience was not so straightforward.

Navigating dimensions between theory and practice has highlighted the fact that we cannot rely on theory alone to inform decision-making. Grounding theoretical ideas within attempts to create real exhibits in specific contexts requires an active dialogue between theory and practice, rather than a one-way conversation.

Design efforts that are led with emotion in mind cannot simply draw from a one-to-one match between a theoretical idea and a corresponding design practice. Emotion is just too complex for this to work—what invokes confusion in one visitor might inspire curiosity in another. Instead, we found more utility in conjecturing broadly about relationships between design and the varied ways emotions might play a role in engagement over time, coupled with iterative testing and assessments of these assumptions.

Relatedly, we found that not all of the project's research studies were equally useful to the design team. Our more complex studies that leveraged high-tech equipment did not always provide designers with insights from the data quickly enough to be useful. Instead, many of our more actionable results emerged, not from more quantitative and subconcious measures like electrodermal activity or other externally observable cues, but from small-scale studies focused on describing visitors' experience in their own terms. We often found that framing our research plans in ways that drew on practical ways of knowing, rather than theory alone, led to more fruitful learning.

With these ideas in mind, we believe that future work addressing emotion in design cannot simply apply the guidelines for practice that we have outlined here and expect straightforward results. Instead, we highlight the need for continued reflection on and testing of how context (social, physical, and environmental) and individual diversity intersect with the design decisions we make.

Negative Can be Positive, for Visitors and Designers!

Early in the project, our team felt disruption as designers and researchers came face-to-face with the fact that these exhibits were, in fact, causing disequilibrium among learners. While interpreted as a success now, we had to navigate our own discomfort with making visitors feel thrown off. Particularly with "Mystery Skulls," team members struggled to acclimate to a new, emotionally inclusive perspective on the impacts of our design decisions. Observing visitors second-guessing themselves at the exhibit—and then reviewing aggregated data that suggested most visitors felt unsure of themselves at distinct points in their engagement with the exhibit—seemed like evidence that we were failing. Team conversations at this juncture illuminated the value of the interdisciplinary nature of the team, as different team members helped put these experiences in perspective: visitors' uncertainty meant we were achieving our goals!

While we have gotten a better sense of what makes struggle productive, we recognize that there are still many avenues to explore about framing negative emotions in a positive light. As discussed previously in this Guide, there are cultural factors at play in how individuals perceive and respond to negative and positive emotion states. While our team worked to address cultural and individual variability in emotional experiences, we acknowledge that there is more work to do to unpack the biases and assumptions inherent in developing exhibits that intentionally foster specific emotional paths. Future research might seek to explore more directly how culture, identity, or other individual characteristics play a role in visitors' productive struggle (or other emotional) experiences. For example, in 2020, the Museum of Science and Ed Together began a new, National Science Foundation-funded project titled *Appraisal in Diverse Populations: Pilot Research About Intersectional Identity in Science Exhibits* (DRL-1906688) that embraces intersectional identities and applies appraisal theory from the affective science literature to shed light on the ways exhibit design contributes to feelings of welcome, belonging, exclusion, and discomfort.

Three Areas for Future Work:

Productive Struggle Beyond Science Museum Exhibits

Our definition of productive struggle leans heavily on the context of science exhibits. Our productive struggle exhibits disrupt visitors through physical challenge, surprising science content, novel phenomena, and challenging questions. Each of our exhibits were task-oriented, asking youth visitors to accomplish goals, although leaving room for more open-ended engagement. But, we do not think productive struggle experiences happen only at science exhibits, through task-oriented experiences alone, or only for learners in the narrow age range of 10 to 17 years old.

Through reflections on this work with professionals outside of the ISE field, **we began to wonder about the potential to design for productive struggle in other contexts**, and with learners outside of our project's targeted age range. Questions we hope to see addressed in future work beyond our walls include:

- How might art museums leverage The Productive Struggle Framework, particularly with exhibitions or gallery themes that surprise, confuse, or even sadden visitors?
- What design strategies are salient in historical or memorial sites, where visitors might confront novel ideas about cultural practices, or might even experience anger or frustration towards societal issues?
- What types of experiences might foster productive struggle in younger children, or older adults?
- How might educator-led programs be developed and facilitated to purposefully lead learners through productive struggle with science topics?

While the content of this Guide outlines the team's ideas about what productive struggle is in the context of science museum exhibits, we acknowledge that other ways of characterizing the productive struggle arc could be equally valid, and supporting design strategies might look quite different based on unique aspects of the environment, context, and social norms of your own site.

Continued Exploration of Emotion in Museum Experiences

As described in Part I of this Guide, productive struggle is not a completely new idea. Other frameworks and theories of meaningful learning also integrate negative and positive feelings, challenge and satisfaction, difficulty and fun. The Productive Struggle Framework contributes to this larger conversation about learning and design and is unique in that emotion is at the center of the work rather than addressed as an additional outcome of meaningful learning experiences. In the Framework, emotional goals and experiences are defined and attended to in design, while research and evaluation explicitly seeks to observe, describe, and make meaning around the emotional experience of learners during the development process.

This project encouraged us to consider what design and development can look like when a focused emotional goal is centered. Acknowledging that there is huge diversity in the emotional landscape of the visitor experience, **our work suggests a broader future trajectory for emotion work in museums**:

- What value might we find in expanding our experience development approaches to consider emotional variability more intentionally?
- What can design look like when emotional goals *other than* productive struggle are forefronted?
- How might future research and development consider design for emotion at the full gallery level?
- How might we more effectively curate emotionally rich–and emotionally integrated–exhibition spaces that offer a variety of emotional experiences across an entire museum visit?
- What opportunities for other trajectories of emotional engagement might best complement productive struggle exhibits, what aspects of design might support such emotional trajectories, and how might these interact to influence visitors' overall museum experience?
- What more do we need to learn about the emotional needs and norms of our visitors to help us best design with diverse audiences in mind and ensure emotional accessibility?

Our team worked together from the beginning of this project to identify ways to capture nuance and complexity in visitors' emotion experiences over time. Combinations of survey items, open-ended questions, stimulated recall interviews, biometric data, and recorded observations helped the team build rich and nuanced stories about visitors' experiences with productive struggle exhibits. In doing this work, we have identified two areas where future research might help inform how we make sense of visitors' emotional experiences in naturalistic museum environments.

Making Sense of Visitor Self-reports:

One strand of inquiry has to do with how visitors talk about their emotional experiences when given the opportunity to narrate in open-ended ways. The youth interviewed for our studies described their feelings in a huge range of ways, sometimes using more obvious language, like "happy" or "motivated." Other times, visitors expressed their feelings using language that seemed just slightly removed from traditional emotion descriptions, such as "mind-blown" or "I felt smart." Still other youth described what they felt in more ambiguous terms, rather than using any emotionally-resonant language at all, such as, "I felt like I really wanted to finish all the challenges," or "I kind of felt like

I was going to get it right." Hearing these responses and working to make sense of them during analysis was quite challenging for the team. We reflected on the nature of emotions, how there are not always clear, objective words that are readily accessible to define one's felt experience, and wondered how to best parse and categorize these diverse narratives. Our approach was to sift carefully through responses, discuss ambiguous language as a group, and refrain from over-interpreting comments that were not clear or contextualized by other data. Future research might examine this phenomenon more closely, exploring how young people tend to explain their felt experiences, and considering how some of these more amorphous feelings might relate to learning experiences.

Sorting through Social Interactions:

Another area of inquiry that remained at the forefront throughout our work was that of social interactions at productive struggle exhibits. While the team worked to describe social interactions and contexts as part of data collection, with one study specifically examining experiences of visitors who engaged with an exhibit alone versus in groups, we continued to grapple with how to represent the complexity of the social elements of emotional experiences within the final framework. Given that social relationships are complex, and that the social context of an experience can be unpredictable in a museum, we had trouble identifying straightforward connections between design features and social elements of visitors' productive struggle experiences (e.g., what fostered disequilibrium through competition for some also prompted collaborative meaning-making for others). However, it was evident that social interactions did play a role in visitors' engagement, as visitors tended to report that doing the exhibits alone or with group members often influenced how they felt. Further, our research with the "Mystery Skulls" exhibit found preliminary evidence that youth engaged with struggle differently if they were with an adult (who tended to ease participants' struggle) compared to a peer (who tended to heighten struggle). New questions emerged for future consideration: What difference does it make if visitors prefer competitive or collaborative experiences? What aspects of design are most salient to visitors engaging solo versus in a group? How can design more intentionally prompt social experiences of productive struggle? How can we create exhibits that provide the appropriate supports for different groups, each with their own social contexts and dynamics? At the time of this writing, the team is still actively investigating whether and how design can support meaningful social interactions at productive struggle exhibits, with an eye toward unearthing the varied pathways of engagement that social interactions might prompt.

Leveraging Emergent Technologies to Advance Emotional Measurement in Design-based Research (DBR)

Our research methods helped us characterize the core features of productive struggle and describe some of the finer grained details of individual productive struggle experiences. This was intensive work, as much of our DBR process was spent iterating–not just the exhibit design, but also the measurement approaches we used. Qualitative data was rich and useful, but took time to analyze. The technology and software used for some data collection and analysis were evolving rapidly during the project's timeline, sometimes making it challenging to gather comparable data consistently. Measures and analysis protocols used earlier in the project did not always make their way to our final studies, meaning we were not always able to make clear comparisons over time. While our flexibility in the DBR process was an asset to informing meaningful and context-specific design, we also acknowledge that future work might require more streamlined measures.

With the growing capabilities of technologies that can detect patterns in engagement and affect over time, we look forward to future research and development initiatives that seek to streamline data processing to support interpretation of fine-grained data in time series. For example, when using the *iMotions* platform it was possible to monitor and sync several streams of engagement data, but we sometimes found ourselves wanting to integrate data streams that were not supported by this software as it was built. In particular, we imagined systems that might track and monitor changes in bodily movement (e.g., the "Sneak" exhibit used a Kinect to measure speed), facial movements (e.g., captured by video cameras), and exhibit use (e.g., the timing of button presses captured by internal logging, such as in the "Mystery Skulls" exhibit) simultaneously to help describe visitors' engagement–but such a software system would have to be built specifically for our context. While the scope of our project did not support development of such systems, our team's experience in collecting and making meaning of rich emotional data streams has laid the groundwork for such research and development inititatives that advance strategies for collecting meaningful emotion data from visitors. We believe such work could be further enhanced by leveraging artificial intelligence (AI) to help make sense of the larger data sets that are inherent when collecting streams of data in time-series.

Beyond *making sense* of data, we also see opportunities to leverage AI to develop responsive or adaptive experiences. Currently our design strategies are implemented in a standardized way across all visitors. While we acknowledge that diverse visitors bring unique needs, interests, and emotions with them, our current research and design capacity is limited to developing experiences that statically offer emotional entry points and supports to visitors as they engage. Though we use UDL guidelines to ensure our exhibits are accessible to broad audiences, we see the potential for leveraging tech-enabled experiences that might more flexibly adapt to the individual emotional needs of visitors. We imagine that such systems could engage visitors in productive struggle experiences in more adaptable and personally relevant ways, using rule-based or machine learning approaches that present different responsive pathways depending on a user's social, physical, or emotional signals. This area of innovation would bring significant challenges, as such technologies and systems–thus far–have primarily been leveraged in more controlled laboratory or classroom based settings (see Bosch et al., 2016; Hutt et al., 2019). However, the promise of such innovation is beginning to unfold, as more informal learning institutions begin to explore the potential for AI and adaptive technologies to expand our conceptions of what it means to build interactive and immersive offerings (see Cieko, 2017).

Finally, we cannot address directions for future work that involve the inclusion of AI-enabled technologies without also acknowledging the complex ethical considerations in this space. Such work runs risks associated with algorithmic bias, or the potential for an algorithm to make incorrect decisions about a user that might hinder, rather than promote, equitable engagement and outcomes (Danks & London, 2017). Many resources are available to help developers mitigate such risks, ranging from research on appropriate data collection and sampling techniques during development of AI systems (see Barocas et al., 2017; Williams et al., 2018), to more targeted tools and support systems, such as those developed by the Algorithmic Justice League (ajlunited.org). More specific ethical considerations arise when dealing with technologies that purport to detect emotions (see Cooney et al., 2018). Our team wants to highlight the importance of informed consent for study participants, the critical role of transparency during the deployment of any system that monitors and collects data from public audiences, and the crucial benefits (for the user and the developer) of being responsive and deliberative when inviting end-user input on such systems during their development (Owen et al., 2013).

Now it's your turn!

Throughout this Guide, we have shared data and other evidence from our work, as well as case studies exploring how our team collaborated across disciplines to develop exhibits that foster productive struggle. We hope these case studies, and our final reflections, illustrate the complexity of doing this work, while also highlighting that it can be manageable and fruitful. As readers of this Guide make moves to begin testing and applying The Productive Struggle Framework in your own settings, remember–we are here to help! Please contact us at evaluation@mos.org to learn more, ask questions, and continue the conversation.

FAQ

After reading this Guide, you may still have questions about how to implement this work at your own institution, about productive struggle, or about emotions in general. Below, we answer some questions that the team imagines other professionals may have about this work. We are still learning about the process of making meaning of emotion in the museum context, and we still have questions too! So, we encourage you to try out different strategies and methods to fulfill your own needs for productive struggle and share what you learn with others.

How much productive struggle do I want in my museum?

In the *What is Productive Struggle?* section, we talk about the concept of "struggle budget." Productive struggle is an active experience that can be mentally and physically taxing. It would be difficult for people to maintain that emotional state during the entirety of a museum visit. Simply put, people may need time to recharge after a productive struggle experience. It is, therefore, important to design for emotional variety within any gallery. As discussed in the *Introduction and Why Emotion?* sections, designing for emotional variety promotes emotional accessibility by making sure that there is something for everyone, no matter what their emotional skills and preferences may be. While our research shows that experiencing productive struggle at exhibits is valuable, educational, memorable, and feels like authentic science to visitors (Todd et al., 2021), there are times when you might have other goals–such as fostering creativity or empathy–for which other emotional trajectories might be more valuable. In short, we recommend applying productive struggle in moderation–when it best supports your goals–and we think it is most effective within the context of a larger exhibition that can support a range of emotional experiences.

How do you decide that productive struggle is a goal for an exhibit component you are developing?

The Invite stage of the framework stresses the importance of making an accessible and welcoming exhibit for all visitors. This stage addresses design elements that can be included in all exhibits, such as providing easy orientation, to achieve this goal. Our framework is designed to support visitors through struggle, but if achieving disequilibrium is turning people away then more support should likely be incorporated. We know from our prior research that visitors are already experiencing negative emotions in museums (see *Why Emotion*?). Our framework is about normalizing struggle and giving visitors the tools to make it productive. Ultimately, we want people to feel challenged, but not so frustrated that they leave. A good productive struggle exhibit should make a museum space feel *more* welcoming, as it widens the range of emotions in that space–hopefully beyond Western assumptions about emotions and preferences for high-energy, positive experiences (Tsai, Knutson, & Fung, 2006). From an equity lens, we seek to invite all visitors to engage with challenging experiences. We uphold the idea that advancing equity involves broadening access to safe, challenging experiences that push our visitors forward and uplift them for who they are.

How do you decide where productive struggle fits in your overall visitor experience?

What is the emotional landscape of your exhibit halls? Are there large areas with high energy exhibits that lack an area for people to recharge? Or are there ample calm areas that might need an emotional kickstart? These questions can help guide your team in figuring out what your emotional needs are in a gallery.

You may already have productive struggle elements at your museum, but you may not have previously considered it that way. Many of the exhibits at the Museum of Science, Boston had aspects of productive struggle before we began this project. We recommend exploring your own exhibits with new eyes, thinking about ways you might strengthen or highlight what is already happening in your galleries, and considering the position each exhibit occupies within your museum's emotional landscape.

I don't have a lot of money! What can I do about productive struggle?

Ultimately, we have found that testing prototypes early and often can reduce the overall costs of exhibit development, because we are able to avoid the costs of fabricating a final exhibit that does not work. The key is to quickly create inexpensive versions of experiences that help you work out the kinks. We frequently test with paper prototypes first, having a live educator read aloud the text that would ultimately be integrated into labels while another team member facilitates changes to the state of the materials (e.g., revealing an answer to a prompt) in ways that will eventually be incorporated into a stand-alone design. When it comes to data collection, we often found the most useful data we gathered about productive struggle came from observations, surveys, and interviews–all of which can be done with minimal materials such as pen and paper. You do not need a research background to solicit user feedback. In fact, it can be especially valuable when the people developing the exhibit experience are able to talk directly to visitors who are trying it out.

You can start with asking people simple questions, like how easy or hard an exhibit was and if that experience was boring, challenging, or frustrating. Then, ask for explanations of visitors' answers. The *How do you Measure PS?* section outlines possible methods you can use, and the *Instrument Appendix* offers additional questions to ask visitors about their productive struggle experiences. We hope that the Framework serves as a helpful logic model and a starting point for designing experiences that contribute to productive struggle and can also help with planning your overall exhibit goals and your prototype development process.

My museum audience is mostly young children! What can I do?

This project focused on youth ages 10-17, which means that there is still a lot to learn about how to support productive struggle among younger age groups and adults! We suspect that people of all ages already experience productive struggle in museums. There is evidence within the developmental and education literature that these kinds of experiences are not only authentic to children's everyday lives, but–when supported–can be valuable and meaningful to them (Vygotsky, 1980; Fischer & Bidell, 2006). Try looking at your exhibits alongside the Framework and doing some low key exhibit testing. Are there components at which young children struggle? Are they able to achieve the exhibit task and leave the exhibit feeling satisfied? If not, what adjustments could you make to ensure this happens? You can adapt the methods we used with youth to serve your own needs–review the *How do We Measure Productive Struggle?* and *Instrument Appendix sections for inspiration*.

Another avenue is thinking about the typical composition of groups who visit your museum. Young children do not visit museums alone, and there may be other group members who would be a good target for productive struggle. For example, could labels be used to disrupt a caregiver's understanding about their child's behavior? There is great potential for future work surrounding productive struggle and young children, and we encourage you to try things out or seek out opportunities to collaborate!

Isn't talking about feelings awkward?

It can be, but our experience has been that visitors are receptive to talking about their emotional experiences at the museum. Inquiring about emotions can offer an opportunity for visitors to talk about experiences they may not have mentioned otherwise, such as how an exhibit relates to memories or prior personal experiences. Asking about emotions may also uncover feedback about visitors' cognitive response to an experience, providing insight on what visitors learned or found valuable in an exhibit. When we only ask visitors what they have learned, they tend to focus on traditional notions of content learning (e.g., facts). When we ask about emotions, visitors have shared deeper reflections on their interests, values, sense of self, and what really matters to them as people.

We asked about emotions in multiple ways during data collection, as we expected some people would be more comfortable with some methods than others. We included open-ended interviews and surveys, as well as emotion word banks or even emojis to select from, as we learned that emotions can be hard for visitors to describe and we wanted to embrace ways to include non-traditional expressions of emotion. Our team also kept a broad definition of "emotions," and accepted whatever response people offered, even if it was a little abstract (e.g., "I feel itchy!")

What does emotion have to do with learning science?

Emotions are present in everything we do, including our decision-making and how we perceive our surroundings. The *Emotion 101* section describes how emotions are tied to our appraisal of different situations, such as if they are positive or negative, and help us decide how to engage with the world. Research has shown that having a broad emotional vocabulary is beneficial to people's emotional regulation skills, which is useful in tackling difficult situations or tasks (Barrett, 2019). We think this is important because science itself is an inherently emotional process in which productive struggle is a normal part of the work. By supporting visitors when they encounter these potentially disruptive emotions, we are helping them safely practice emotional regulation skills that they can take with them outside of the museum. Even learning about the term "productive struggle" can also help empower visitors to put words to this complex emotional arc!

Can productive struggle be unwelcoming or turn people away?

The Invite stage of the framework stresses the importance of making an accessible and welcoming exhibit for all visitors. This stage addresses design elements that can be included in all exhibits, such as providing easy orientation, to achieve this goal. Our framework is designed to support visitors through struggle, but if achieving disequilibrium is turning people away then more support should likely be incorporated. We know from our prior research that visitors are already experiencing negative emotions in museums (see Why Emotion?). Our framework is about normalizing struggle and giving visitors the tools to make it productive. Ultimately, we want people to feel challenged, but not so frustrated that they leave. A good productive struggle exhibit should make a museum space

feel more welcoming, as it widens the range of emotions in that space–hopefully beyond Western assumptions about emotions and preferences for high-energy, positive experiences (Tsai, Knutson, & Fung, 2006). From an equity lens, we seek to invite all visitors to engage with challenging experiences. We uphold the idea that advancing equity involves broadening access to safe, challenging experiences that push our visitors forward and uplift them for who they are.

Can you design for productive struggle in programmatic experiences?

We believe that the answer is a resounding "Yes!" Although this project focused on creating exhibits, we see expansion of the framework to programs as an intriguing next step. In fact, a number of teachers in formal education settings are already using the term "productive struggle," especially in conjunction with mathematics topics (see *Is Productive Struggle a New Idea?*)

Isn't this manipulating people?

As designers, we are always affecting people's emotions. Why not be intentional about encouraging emotions that support our goals? Struggling through a task is an important part of both science and learning in general. Although the framework helps us purposefully design for struggle, it also helps us make sure we embed design features that support visitors in multiple ways–cognitively, socially, and physically–so visitors can leverage our experiences for learning in a way that suits them. Further, our research found that visitors experience a wide range of emotions at productive struggle exhibits, not just those outlined in the framework. In fact, many framework strategies emphasize offering multiple paths for people to find what is comfortable for them. Lastly, we found that, when given the opportunity, people generally do try the harder, more challenging options. And although our designs are pushing them towards struggle, our aim is to ensure visitors feel free to decide how much or how little they wish to engage with it.

Glossary

This glossary contains definitions of keywords and concepts as utilized by the Productive Struggle team.

Active Prolonged Engagement
Exhibit experiences where museum visitors lead their own learning, have extended dwell times, and show variety in their exhibit interactions. Visitors' actions build upon previous actions within the exhibit (Humphrey & Gutwill, 2017).

Activation/Arousal
The amount of energy associated with a feeling. For example, heightened physiological activity might be described as feeling very active, or lower activity could be described as lethargic.

Affective Science
An interdisciplinary field that draws from neuroscience, psychology, health, computer science, economics, anthropology, and more to investigate the nature of emotion and its role in society.

Appraisal
Ongoing, largely unconscious judgments about whether an experience is: good or bad; relevant or irrelevant to one's goals; comfortable or threatening; novel or familiar; within or outside of one's control; and consistent with or opposed to social norms. Appraisals are associated with bodily changes in heart rate, blood pressure, and respiration.

Biometrics
A type of data collection that involves measuring aspects of the human body using technology. Examples of this type of data can include eye-tracking, electrodermal activity, and expression analysis.

Cognitive Dissonance
The mental conflict experienced by a learner resulting from inconsistencies within their "knowledge, thought[s], or belief[s]" (Festinger, 1962, p. 3). This conflict arises from people's natural desire to resolve these inconsistencies. Resolving cognitive dissonance can be associated with changing behaviors, opinions, or attitudes.

Core Affect
The basic sense of how one's body feels that can be described by valence and arousal. It is distinct from emotion which is a complex mental construction.

Desirable Difficulty
A challenging learning experience that makes learners' brains process information in ways that are more memorable (Bjork & Bjork, 2011).

Disequilibrium
A sense of imbalance that can be experienced as emotions like confusion, frustration, surprise, or unease.

Electrodermal Activity
Biometric measurement of activation level based upon measuring electrical conductivity of skin. These readings change depending on how calm or active one feels, which relates to changes in the amount of minute quantities of sweat created by the learner's body. Pair electrodermal activity with self-report data (e.g., storyboarding, stimulated recall, surveys) so participants can add meaning that helps with interpretation of this data. Also called "skin conductance" and "galvanic skin response" (GSR).

Emotion
Biological, cognitive, and psychological processes that result in the experience of feeling that aligns with a cultural construct for a specific state (e.g., anger, happiness, etc.). Emotion impacts thoughts and behavior.

Eye-tracking
Biometric measurement of where someone is looking that can provide information about visitors' attention during an experience. This type of data can be used to understand what participants looked at, how long they looked, and how long it took someone to look at something from the start of the experience. Eye-tracking technologies also allowed the Productive Struggle team to capture videos from the visitor's point of view.

Flow
The feeling of being "in-sync" and in balance with one's work or tasks. This concept involves being fully absorbed by and being intrinsically motivated by activities which are at appropriate levels of difficulty. Learners draw upon current skills and extend them while feeling a sense of emotional balance (Nakamura & Csikszentmihalyi, 2002).

Gestalt
A sense of how one feels taken as a reflection on a whole experience. This stands in contrast to a momentary assessment of feeling.

Hard Fun
When a learner chooses to engage in a challenging activity and finds it enjoyable. In this experience, learners knowingly seek out and select the activity because it is difficult and enjoy the activity in the moment (Papert, 2002).

Maker-centered Learning
An instructional framework developed by Agency by Design, a research initiative out of Project Zero. It centers on three critical maker capacities that support a sensitivity to design, which in turn encourages a sense of maker empowerment: (1) looking closely; (2) exploring complexity; and (3) finding opportunity.

Productive Failure
A learning theory that recommends designing learning experiences with little scaffolding that enable learners to experience short-term failures–while avoiding frustration–but show increases in long-term understanding (Kapur, 2008).

Productive Struggle
An experience with three elements: 1) a learner encounters a challenging task and feels disequilibrium, which might be experienced as emotions like confusion, frustration, surprise or unease; 2) the learner is supported to engage with and persist in the task; and 3) the learner achieves a positive resolution, which might be experienced as emotions like satisfaction or pride.

Productive Struggle (Formal Mathematics Education)
A term originating from formal mathematics education to describe intentional efforts to design classroom activities that support learners through struggle towards a productive resolution (Warshauer, 2014).

Social and Emotional Intelligence
Our capacity to understand, use, and manage emotion. Social and emotional intelligence involves skills like self-management (e.g., controlling impulses and managing stress and goal setting), social awareness (e.g., perspective taking and empathy), relationship skills (e.g., cooperating, collaborating, and active listening), and decision making (e.g., analyzing and solving problems, reflecting, and ethical responsibility).

Stimulated Recall
An interview technique used to understand visitors' emotional experience in using an exhibit. A participant narrates what they were feeling during their experience using an exhibit as they watch a video recording of themselves at the exhibit and respond to researchers' questions about specific events of interest.

Storyboarding
An interview method used to understand visitors' emotional experience at exhibits. After using an exhibit, a participant is given a series of cards asking them to describe what they did at the beginning, middle, and end of using the exhibit. Then, they are asked to describe how they felt at each stage. Researchers ask questions to probe about visitors' emotional experiences.

Struggle Budget
The concept that people's bodies have a limited physiological capacity to be in an emotional state of disequilibrium. In design, attending to a struggle budget means using clear exhibit design strategies to minimize the burden of figuring out usability and helping visitors focus their intention on the designed challenges.

Subjective Feeling
Commonly called emotion, this is a concept that constitutes one's understanding of their overall experience, which we label with familiar emotion terms like happiness, sadness, rage, pride, or relief.

Universal Design
The philosophy of designing for extreme use cases, not the average. By doing this environments and products become more usable for everyone.

Valence
The sense of how positive or negative a feeling is, from unpleasant to pleasant.

Zone of Proximal Development (ZPD)
The space between what a learner is currently able to do on their own and what they can achieve with assistance from a more knowledgeable peer, teacher, or other person. The types of learning that are within the ZPD for a given learner changes over time, and the "length" of the ZPD or what it includes differs for each learner (Vygotsky, 1980).

Instrument Appendix

The example instruments outlined below, while created for this Guide, are drawn from real instruments utilized during formative evaluation of the productive struggle exhibits. Additional information outlining the purpose and use of each question can be found in the section *How Do We Measure Productive Struggle?*

Example Formative Observation Form

Below is a basic observation form used during formative exhibit testing. During our testing we created sections specifically for behaviors related to productive struggle, but you may decide to have an open notes page instead. The observation checklist below is only an example of some of the items we used, so not all of these will make sense for every exhibit.

Evaluator Initials: ___ Date: ___ Group #: ___ # Adults: ___ # Kids: ___

Time start:_____

Did the visitor use:

❑ Exhibit element 1
# of attempts:_____

❑ Exhibit element 2
# of attempts:_____

❑ Exhibit element 3
# of attempts:_____

Usability/Accessibility Issues (check off and explain below)

❑ Confusion with instructions
❑ Confusion about content
❑ Other usability/accessibility issues

Usability, accessibility, or undesirable confusion issue
(Exhibit is broken or visitor is unable to use exhibit, etc.)

Evidence of disequilibrium	Evidence of persistence	Evidence of productivity

Time end: _____

Example Survey

Did you feel any of the following emotions when you were using this exhibit? Circle any emotions you felt:

Confused	Confident	Satisfied	Neutral
Frustrated	Focused	Proud	Relaxed
Challenged	Motivated	Happy	Comfortable
Bored	Pessimistic	Disappointed	Tired

Emotion phase categories

Disequilibrium: confused, frustrated, challenged
Persistence: confident, focused, motivated
Productivity: satisfied, proud, happy

Emotions not listed within these categories (ex. neutral, relaxed, comfortable, tired, bored, pessimistic, and disappointed) are included to capture a range of possible emotions and may help provide evidence, or lack thereof, for productive struggle.

Example Survey Continued

Which statement best represents how you felt at this exhibit?

- ❑ It felt **easy**, and it was **boring** to keep trying.
- ❑ It felt **easy**, but it was **satisfying** to keep trying.
- ❑ It felt **challenging**, but it was **satisfying** to keep trying.
- ❑ It felt **challenging**, and it was **frustrating** to keep trying.

How much do you agree or disagree with these statements?

	Strongly disagree	Disagree	Agree	Strongly agree
It was easy to figure out **what to do** at this exhibit.	❑	❑	❑	❑
It was easy to figure out **how to use** the exhibit.	❑	❑	❑	❑
The exhibit seemed **broken**.	❑	❑	❑	❑

This question was left out when asking about exhibit design features as this is included as one of the choices

Design Features Survey Questions

These survey questions were asked if participants indicated that they felt productive struggle emotions from the question, "Did you feel any of the following emotions when you were using this exhibit?" and if they answered with feeling challenged and/or satisfied from the question, "Which statement best represents how you felt at this exhibit?". We used branching logic on our online survey platform to only display the appropriate questions and to fill in the question wording with the words the participants had selected. Depending on what aspect of the framework you are testing and what features are present for your exhibit, you would only include the relevant options for each question. For more information see the part titled "Linking design and emotion" in the *How Do We Measure Productive Struggle?* section.

1. You said you felt [disequilibrium emotion(s)]. What made you feel that way?

- ❑ Figuring out how to use the activity.
- ❑ Figuring out what the activity was about.
- ❑ The activity seemed broken.
- ❑ Having to guess the animal right away. (Mystery Skulls)
- ❑ Answering questions about the skull features. (Mystery Skulls)
- ❑ Guessing what the animal was at the end. (Mystery Skulls)
- ❑ Not having enough information to make decisions. (Sneak, Mystery Skulls)
- ❑ Having other people at the activity.
- ❑ Doing the activity by myself.
- ❑ Figuring out how the air was moving. (Air)
- ❑ Using the air to get the ball to do what I wanted. (Air)
- ❑ Figuring out how to sneak up on the bird (Sneak)
- ❑ Moving slowly to sneak up on the bird. (Sneak)
- ❑ Other (describe): _____

2. You said you felt [persistence emotion(s)]. What made you feel that way?

- ❑ I could make choices about what to do.
- ❑ I wanted to get the right answers. (Mystery Skulls)
- ❑ I wanted to learn more.
- ❑ I made a mistake and I wanted to do better.
- ❑ I liked putting the clues together to solve the mystery. (Mystery Skulls)
- ❑ The exhibit helped me break things down piece by piece. (Mystery Skulls, Air)
- ❑ I could do the activity at my own pace. (Mystery Skulls, Air)
- ❑ I wanted to finish what I was working on.
- ❑ I wanted to do all the challenges. (Air)

- ❑ I wanted to do both of the birds. (Sneak)
- ❑ I wanted to do all the skulls. (Mystery Skulls)
- ❑ Doing the activity by myself.
- ❑ Having other people at the activity.
- ❑ It was hard but it felt ok to keep trying.
- ❑ There were multiple ways to do the activity.
- ❑ The activity told me to try again. (Sneak, Mystery Skulls)
- ❑ I could tell I was making progress.
- ❑ I wanted to solve the challenges.
- ❑ I liked using what I learned about how air works to solve challenges. (Air)
- ❑ I could try it multiple times.
- ❑ Other (describe): _____

3. You said you felt [productivity emotion(s)]. What made you feel that way?

- ❑ Learning new information.
- ❑ Putting information together to solve the skull mystery. (Mystery Skulls)
- ❑ Using what I learned about air to solve a challenge. (Air)
- ❑ Seeing the deer at the end. (Sneak)
- ❑ The look and feel of the activity.
- ❑ There were things to touch and interact with.
- ❑ Getting a question right. (Mystery Skulls)
- ❑ Being in the green zone of the Sneak-o-meter. (Sneak)
- ❑ Hearing the bird songs. (Sneak)
- ❑ Seeing the bird respond to my motion. (Sneak)
- ❑ I got the ball to do what I wanted it to do. (Air)
- ❑ Getting more information after getting a question wrong. (Mystery Skulls)
- ❑ Having other people at the activity.
- ❑ Doing the activity by myself.
- ❑ Moving slowly to sneak up on the bird. (Sneak)
- ❑ Learning a new skill.
- ❑ Other (describe): _____

Example Interview

Below are some of the typical questions we asked during formative testing of our exhibits. Many of these questions depend on visitors' survey responses, which means the length of the interview can vary from individual to individual. For ease of use, we have indicated which survey response relates to each interview question using brackets at the beginning of the sentence. For example "[1]" means this is dependent on the question "Did you feel any of the following emotions when you were using this exhibit?" If a bracketed number is not present, then the question was asked of all participants.

I1. [1] On your survey, you said you felt [emotion word]. Can you explain why you felt that way?

I2. [2] On your survey you said the exhibit was [easy/challenging] and [boring/satisfying/frustrating]. Could you explain why you felt that way?

I3. Why did you decide to leave the exhibit when you did?

I4. What, if anything, did you learn from this exhibit?

Our team only asked when the emotions were related to the phases of productive struggle, but you are welcome to ask for emotions that feel relevant to your needs."

I5. [3] On your survey you said it was hard to figure out what to do at the activity. Can you explain why you felt that way?

I6. [3] On your survey, you said it was hard to figure out how to use the activity. Can you explain why you felt that way?

I7. [3] On your survey, you said the activity seemed broken. Can you explain why you felt that way?

I8. What parts of the exhibit worked well for you? (If they have difficulty answering: As we make changes to the exhibit, what should we make sure we keep the same?)

I9. What about this exhibit was confusing or hard to use?

I10. Anything else?

Storyboarding

Below are the cards we used for storyboarding (see "Guided Recall Activities" in the *How Do We Measure Productive Struggle?* section) as well as instruction language from our protocol for how we would introduce the activity to visitors. Visitors were allowed to use as many "next" cards as needed. Feel free to make copies and use them yourself or adapt them to your own needs.

Protocol Language

"In a second we'll make a storyboard about how you felt at this exhibit. This is not a test– we're just trying to understand what your experience was like. As you were at the exhibit, I jotted down some notes about what you were doing. [Show cards and read aloud]: So first, you [...] and then you [...], and right at the end you [...].

Is that about how you remember what you did at the exhibit, or is there anything you'd like to add or change? [adjust cards as appropriate based on response].

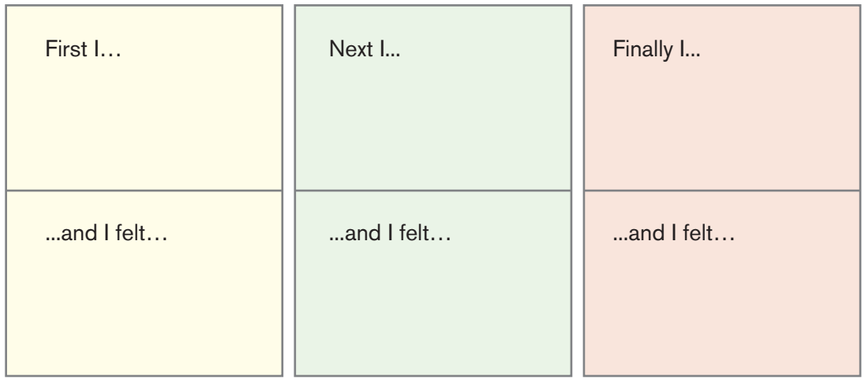
Great, thanks! Now we're interested in how you felt while you were using the exhibit, and why. When we talk about feeling, it might be a specific emotion like happy or angry, or it might be more reflective, like confused or confident. There are no right or wrong answers! So for each of these cards, could you write in or tell me how you were feeling at those points? I have these stickers with different feeling words that you can put on the cards, or you can come up with your own words and we can write them in."

As appropriate, probe from your observation notes with questions such as:

- When you were [observation behavior], I heard you say, "[quote]." What were you feeling when you said that?
- I noticed you [observation behavior]. What were you feeling when you did that?
- The exhibit [told you to try again, etc.]. What were you feeling when it did that?

Other additional questions to ask include:

- Why do you think you felt [emotion] when you [action]?
- Did you feel any other emotions at this point?
- Was there anything else you felt that we haven't put on this storyboard?



Stimulated Recall

Other than the video equipment to record visitor behaviors and responses, we only used a plain piece of paper for taking notes. Like storyboarding, we have provided our protocol language which outlines the activity as well as the types of questions we asked below.

Protocol Language

"Now I'm going to play the video we took while you were doing the activity. As you watch the video, I'd like you to describe to me out loud how you're feeling and why. So you might say things like 'I'm confused about what kind of skull it is,' or 'I'm laughing because I made a mistake.' I'll be pausing the video at certain points so you can tell me in more detail about what you were feeling, and I'll be taking notes about what you tell me. Do you have any questions? Great. I'm going to start the video now. When it starts playing, you can start describing to me how you were feeling." [Start screen recording].

As the visitor narrates the video, write down emotions and corresponding thoughts or events.

Here are some general tips to keep in mind as you do this protocol:

- If they are not being particularly talkative or descriptive when narrating the video, you can pause it every 30 seconds or so and ask *"Can you tell me about how you were feeling there? Why do you think you felt that way?"*
- Pause the video at points where they're going into detail about a particular event, and follow up those descriptions with questions about what they were feeling - *"Can you tell me about how you were feeling there? Why do you think you felt that way?"*
- If they don't talk about their feelings at time points that you noted on the observation sheet, pause the video at those points and ask *"Can you tell me about how you were feeling there? Why do you think you felt that way?"*

If needed, you can abbreviate stimulated recall by skipping over inactive periods (for example, if not much is changing in their exhibit interaction or description) or playing the video at faster than real-time. A good rule of thumb is to shorten any exhibit interactions that took longer than 5 minutes. Try to keep the entire protocol under 30 minutes.

References

Barocas, S., Bradley, E., Honavar, V., & Provost, F. (2017). Big data, data science, and civil rights. *Computing Community Consortium*. Retrieved from https://pennstate.pure.elsevier.com/en/publications/big-data-data-science-and-civil-rights

Barrett, L. F. (2017). *How emotions are made: The secret life of the brain.* Houghton Mifflin Harcourt.

Bevan, B., Gutwill, J. P., Petrich, M., & Wilkinson, K. (2015). Learning through STEM-rich tinkering: Findings from a jointly negotiated research project taken up in practice. *Science Education, 99*(1), 98-120.

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.

Bosch, N., D'Mello, S. K., Ocumpaugh, J., Baker, R. S., & Shute, V. (2016). Using video to automatically detect learner affect in computer-enabled classrooms. *ACM Transactions on Interactive Intelligent Systems (TiiS), 6*(2), 17.

Brackett, M. (2019). Permission to feel: *Unlocking the power of emotions to help our kids, ourselves, and our society thrive.* Celadon Books.

CAST. (2018). Universal Design for Learning Guidelines version 2.2. http://udlguidelines.cast.org

Clapp, E. P., Ross, J., Ryan, J. O., & Tishman, S. (2016). *Maker-centered learning: Empowering young people to shape their worlds. John Wiley & Sons.*

Cieko, B. (2017). Examining the impact of artificial intelligence in museums. *Museums and the Web*. Retrieved from https://mw17.mwconf.org/paper/exploring-artificialintelligence-in-museums

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *The Journal of the Learning Sciences*, 13(1), 15-42.

Cooney, M., Pashami, S., Sant'Anna, A., Fan, Y., & Nowaczyk, S. (2018, April). Pitfalls of Affective Computing: How can the automatic visual communication of emotions lead to harm, and what can be done to mitigate such risks. In *Companion Proceedings of the The Web Conference 2018* (pp. 1563-1566).

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, 29, 153-170.

Danks, D., & London, A. J. (2017). Algorithmic bias in autonomous systems. In C. Sierra (Ed.) *Proceedings of the International Joint Conferences on Artificial Intelligence* (pp. 4691-4697).

Dancstep, T., & Sindorf, L. (2016). *Exhibit designs for girls' engagement*: A guide to the EDGE design attributes. San Francisco: Exploratorium.

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

Friedman, A. (Ed.). (2008). *Framework for evaluating impacts of informal science education projects*. National Science Foundation.

González-Espada, W. J., Birriel, J., & Birriel, I. (2010). Discrepant events: A challenge to students' intuition. *The Physics Teacher, 48*(8), 508-511.

Granberg, C. (2016). Discovering and addressing errors during mathematics problem-solving–A productive struggle?. *The Journal of Mathematical Behavior, 42*, 33-48.

Hedman, E. B. (2014). *Thick psychophysiology for empathic design* [Doctoral dissertation]. Massachusetts Institute of Technology.

Humphrey, T., & Gutwill, J. P. (2017). *Fostering active prolonged engagement: The art of creating APE exhibits*. Routledge.

Hutt, S., Grafsgaard, J., & D'Mello, S. K. (2019). Time to scale: Generalizable affect detection for tens of thousands of students across an entire school year. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (CHI 2019)* (pp. 476). New York, NY: ACM.

Immordino-Yang, M. H. (2015). *Emotions, learning, and the brain: Exploring the educational implications of affective neuroscience*. W.W. Norton & Company.

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

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences, 21*(1), 45-83.

Kollmann, E. K. (2007). The effect of broken exhibits on the experiences of visitors at a science museum. *Visitor Studies, 10*(2), 178-191.

Linn, M. C., Chang, H. Y., Chiu, J., Zhang, H., & McElhaney, K. (2010). Can desirable difficulties overcome deceptive clarity in scientific visualizations. In A.S. Benjamin (Ed.) *Successful remembering and successful forgetting: A Festschrift in honor of Robert A. Bjork* (pp. 239-262). Routledge.

Lin-Siegler, X., Ahn, J. N., Chen, J., Fang, F. F. A., & Luna-Lucero, M. (2016). Even Einstein struggled: Effects of learning about great scientists' struggles on high school students' motivation to learn science. *Journal of Educational Psychology, 108*(3), 314-328.

Longfield, J. (2009). Discrepant teaching events: Using an inquiry stance to address students' misconceptions. *International Journal of Teaching and Learning in Higher Education, 21*(2), 266.

Lynch, S. D., Hunt, J. H., & Lewis, K. E. (2018). Productive struggle for all: Differentiated instruction. Mathematics Teaching in the Middle School, *23*(4), 194-201.

Majewski, J. (1996). *Smithsonian guidelines for accessible exhibition design*. Smithsonian Accessibility Program: Washington, D.C. https://www.si.edu/Accessibility/SGAED

Martin, L. (2015). The promise of the maker movement for education. *Journal of Pre-College Engineering Education Research (J-PEER), 5*(1), 4.

Museum of Science. (2014). *Universal Design plan for exhibit design & development*. https://www.mos.org/sites/dev-elvis.mos.org/files/docs/misc/MOS_UD_Plan.pdf

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.

National Research Council. (2009). Learning science in informal environments: People, places, and pursuits. National Academies Press.

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

Posey, A. (2018). *Engage the brain*: How to design for learning that taps into the power of emotion. ASCD.

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

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.

Todd, K., Rappolt-Schlichtmann, G., Daley, S., May, S., Paneto, S. (under review). *The importance of designing for productive struggle in museums: Outcomes on visitors' engagement, learning, and perceived value.* Manuscript submitted for publication.

Tsai, J.L., Knutson, B., & Fung, H.H. (2006). Cultural variation in affect valuation. *Journal of Personality and Social Psychology, 90*(2), 288-307.

Tyng, C. M., Amin, H. U., Saad, M. N., & Malik, A. S. (2017). The influences of emotion on learning and memory. *Frontiers in psychology, 8,* 1454.

VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. B. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction, 21*(3), 209-249.

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

Williams, B. A., Brooks, C. F., & Shmargad, Y. (2018). How algorithms discriminate based on data they lack: Challenges, solutions, and policy implications. *Journal of Information Policy, 8,* 78-115.