



The Many Shapes of Formative Evaluation in Exhibition Development

Cathy Sigmond

Creating successful exhibitions does not happen overnight. As anyone who has been part of developing an exhibition knows, designers, curators, educators, and many others contribute their time and expertise. Evaluation adds another critical voice to the process: the visitor's.

Evaluation is about studying something to determine its feasibility or effectiveness, and it is a dynamic part of creating an exhibition.¹ There are three phases of evaluation in exhibition development – front-end, formative, and summative – all of which help museum professionals understand what aspects of an exhibition are (or are not) “working” for visitors. The first phase, front-end evaluation, involves talking to visitors about their knowledge, questions, expectations, and concerns regarding an exhibition's topic or theme. It takes place early in exhibition development, before materials are developed; it is key for providing focus and direction early on in the process.² The last phase, summative evaluation, happens after an exhibition has been installed and helps staff understand how visitors ultimately experience the exhibition and whether that aligns with staff's intentions. Both of these phases of evaluation are important for understanding visitors' potential and actual experiences with an exhibition. But it is the middle phase, formative evaluation, that is really at the heart of exhibition development.

Formative evaluation happens during design development, when the exhibition is more than just a concept (i.e., you have developed some materials), but nothing is final. The goal is to systematically test actual materials and

interpretive strategies in order to make changes to improve them for the final exhibition.³ This timing is why formative evaluation is so crucial. It gives you a chance to understand how what you've developed is actually working for visitors, before it is too late to make changes.

As an evaluator, formative evaluation is by far my favorite type of evaluation. What intrigues me most is its rich variety. Formative evaluation can mean anything from assessing the look and feel of graphics, to understanding how labels convey key messages, to testing usability in digital interactives. All of this is in service to understanding how visitors actively make meaning, or find significance, from an exhibition – whether emotionally, intellectually, spiritually, or in some other way – through a constant process of making connections; and, the extent to which this aligns with a museum's intentions for an exhibition.⁴ In this way, formative evaluation is both exciting and practical.

However, formative evaluation can also feel daunting. When there are so many pieces to an exhibition – and thus many possibilities for what to test – where do you start? What is the best approach to use?

In this article, I begin by identifying principles that underlie all formative evaluations. Then, I articulate various approaches to using formative evaluation for creating exhibitions.

1 Stephanie Downey, “Visitor-Centered Exhibition Development,” *Exhibitionist* 21, no. 2 (Spring 2002), <https://rka-learnwithus.com/wp-content/uploads/2017/10/exhibitionist.pdf>.

2 “Glossary of Visitor Studies Terms,” *Visitor Studies Association*, accessed December 10, 2018, <https://www.visitorstudies.org/glossary-of-terms#e>.

3 Ibid.

4 Lois Silverman, “Visitor Meaning-Making in Museums for a New Age,” *Curator* 38, no. 3 (1995), <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.2151-6952.1995.tb01052.x>.

Every research or evaluation study should start with “why,” and formative evaluation is no exception. What are the core questions you hope to answer through formative evaluation?

Finally, I consider how formative evaluation intersects with two closely-related areas of practice – user experience (UX) and information architecture – and how we might draw inspiration from them. My goals are to 1) shed light on what I see as formative evaluation’s biggest strength: its practical use in many contexts, 2) ease potential anxiety about selecting the best approach by providing useful guidelines, and 3) consider how we might push our thinking on approaches to and possibilities of formative evaluation.

Formative Evaluation in Brief

Chances are you’ve done formative evaluation at some point, whether or not you have called it by its name. In essence, it is about testing something to find and correct problems. A more formal definition, from renowned evaluator and psychologist Chandler Screven, applies this principle specifically to exhibition development: “Formative evaluation,” he writes, “provides information about visitors’ reactions to temporary versions of the most important [exhibits] in terms of both their ability to generate and focus visitor attention and effort, and their ability to ‘deliver’ (communicate) their messages.”⁵

5 Dr. Chandler Screven is Professor Emeritus of Psychology at the University of Wisconsin-Milwaukee. He has been a major contributor to the field of visitor studies for over 40 years. See C. G. Screven, “Uses of Evaluation Before, During, and After Exhibit Design,” *ILVS Review: A Journal of Visitor Behaviour* 1, no. 2 (1990): 41.

This is not a new idea in museum exhibition development or evaluation practice. Those involved in creating exhibitions have been testing their ideas for a long time, and there is a body of published formative evaluation studies as evidence.⁶

Over the years, many people have contributed to our understanding of formative evaluation.⁷ Collectively, they have articulated a few core principles that underlie all formative testing. First, formative evaluation is informal. That is, it is designed to provide feedback in a short time frame and thus lacks the formality of a full-scale research study (though it is still systematic).⁸ Second, it is iterative. Ideally, designs are tested repeatedly until goals are reached. And third, it is varied and flexible. In practice, formative studies look very different depending on goals and available resources. The third principle in particular is worth unpacking, as the beauty of formative evaluation lies in its variety and flexibility.⁹

6 As of September 5, 2018, there were 168 formative evaluation reports published on InformalScience.org (a commonly-used platform for sharing evaluation reports), 56 of which were for exhibitions. Other sources for formative evaluation reports include the Museums and the Web conference proceedings (www.museweb.net/bibliography) and a variety of evaluation journals, including *Visitor Studies* (published by the Visitor Studies Association) and *New Directions in Evaluation* (published by the American Evaluation Association). While it’s clear that formative evaluations are numerous, it is difficult to gauge the true number conducted in museums, as many go unpublished or are not shared broadly since their usefulness is limited to the organization that conducted the evaluation. “Search Results for ‘Evaluation’ and ‘Formative’ on InformalScience.org,” accessed September 5, 2018, [www.informalscience.org/search-results?f\[o\]=search_api_combined_2:41&f\[1\]=search_api_combined_2:11](http://www.informalscience.org/search-results?f[o]=search_api_combined_2:41&f[1]=search_api_combined_2:11).

7 Here are four short, easy-to-read foundational works on using formative evaluation in exhibition development: 1) Samuel Taylor, *Try It! Improving Exhibits Through Formative Evaluation* (Washington, DC: Association of Science and Technology Centers, 1991); 2) Randi Korn, “Studying Your Visitors: Where to Begin,” *History News* 49, no. 2 (1994); 3) Stephanie Downey, “Visitor-Centered Exhibition Development,” *Exhibitionist* 21, no. 2 (Spring 2002); 4) Judy Diamond, Jessica J. Luke, and Dave Uttal, *Practical Evaluation Guide: Tools for Museums and Other Informal Education Settings* (Altamira Press, 2009).

8 Korn, “Studying Your Visitors.”

9 Meghan Stockdale and Elizabeth Bolander, “Mastering the Art and Science of Formative Evaluation in Art Museums,” *Museums and the Web Conference Papers 2015*, accessed September 5, 2018, <https://mw2015.museumsandtheweb.com/paper/mastering-the-art-and-science-of-formative-evaluation-in-art-museums/>.

One Function, Many Forms

Knowing that you can essentially test anything through formative evaluation is exciting, but can also feel intimidating. It helps to compartmentalize. Although there are many ways to break down formative evaluation from a practical standpoint, I've listed several that are particularly relevant to exhibitions. Broadly, they fit into three categories: why (the questions that drive testing); what (the materials to test); and how (the methods to test these questions and materials).

Why: Questions That Drive Testing

Every research or evaluation study should start with “why,” and formative evaluation is no exception. What are the core questions you hope to answer through formative evaluation? Or, from another angle, what are the core aspects of your design that you want to test? Here are some of the most common ideas and questions that are useful for exhibition developers and designers to investigate through formative evaluation:

Visuals: How do the exhibition’s visual elements affect mood and meaning-making? Do they catch visitors’ attention? Spark intrigue and curiosity? Cause confusion?

Usability: This is about an interface’s ease-of-use, whether analog or digital. When visitors encounter your designs, do they know what to do? How easy is it for them to figure it out?

Language: This involves testing word choice and tone. Are there particular words or phrases that visitors find difficult to understand, or that convey a different feeling than intended? To what extent does the text convey key messages?

Organization: This is the structure of information provided. Is it organized in a cohesive way that supports visitors in making sense of the ideas or issues raised? Are there gaps or inconsistencies that prove confusing?

Affect: How do visitors feel after encountering your designs? Did they enjoy them? Are they left with negative emotions?

Relevance: This is about understanding whether something appeals to visitors on a personal level. What, if anything, resonates? Why does it resonate?

A combination: While they can be tested in isolation, all of these concepts together enable or inhibit meaning-making. You may want to study more than one of them, either within a single exhibit component or across many components.

What: The Materials to Test

The prototypes, or preliminary materials, that you show visitors to test these concepts can take many different forms. This depends on a variety of factors, including how far along you are in exhibition development and your resources, such as time, budget, and staffing. Prototypes should be low-cost and do not need to be polished, as your design will change based on what you learn.¹⁰ That said, even though they are rough, you should have clear goals for each prototype and know how those goals relate to the overall goals for the exhibition. Here are some of the most common and effective prototypes to show visitors:

Draft signage: These range from early drafts printed on standard 8.5” x 11”

10 Stockdale and Bolander, “Mastering the Art and Science of Formative Evaluation in Art Museums.”

paper to larger, more refined drafts that include designed elements; for instance, designer-generated labels mounted on foamcore. They are especially useful for testing questions of language, graphics, and organization.

Static prototypes: These are typically a paper version of an interactive; they lack functionality but simulate a general experience for visitors. For instance, you might mock up different screens visitors would see in a touchscreen interactive. Paper prototypes can be used to test all concepts, but are especially useful for testing organization.

Interactive prototypes: These are working (but still unfinished) experiences that closely model the final intended design. For instance, you might work with a media designer to create a working (yet unrefined) version of a touchscreen interactive to show visitors on an iPad®. Or, you might work with your exhibits staff to create a rough mechanical interactive out of cardboard that visitors can manipulate to get a sense of the intended experience. These are useful for testing several concepts, usability in particular.

Rough video or audio: This is an early version of any video or audio you are producing as part of an exhibition. Testing a crude version can help determine if the chosen language, music, and graphics support meaning-making. For instance, does a video spark intrigue or curiosity, or cause confusion? Can visitors find something in a video they relate to personally? How does hearing particular sounds make visitors feel, and is this in line with what you intended?

Methods used in formative evaluation are usually qualitative and can be done quickly with a small sample of visitors – typically, between 10 to 20 visitors. They encourage an immediate feedback loop between developers and evaluators, which allows you to test alternative designs as you go.

A combination: Consider testing a set of prototypes rather than one prototype in isolation. This helps visitors more easily picture the future exhibition and allows you to test how different components work together.

How: Methods to Test These Questions and Materials

The third consideration for formative evaluation is methodology. Which methods should you choose to test your concepts and materials? Again, there are many options. Methods used in formative evaluation are usually qualitative and can be done quickly with a small sample of visitors – typically, between 10 to 20 visitors. They encourage an immediate feedback loop between developers and evaluators, which allows you to test alternative designs as you go. Here are some of the most common methods¹¹ used to gather data in formative evaluations (though they are not mutually exclusive):

A/B Testing: Simply put, this means testing two versions of something at random. The goal is to see which design is preferred and/or easier to use or understand. This method allows for quick testing of virtually any type of exhibition component (e.g., introductory text or

¹¹ Downey, “Visitor-Centered Exhibition Development,” 42.

attract screens for a digital interactive), alternating at random which version visitors see or use first.

Think-aloud protocols: In this method, a visitor is asked to do something while speaking their thoughts aloud. The researcher might ask questions to clarify what the visitor said or probe them to say more; but, it is not a formal interview. Think-aloud protocols are especially useful for testing questions of process, usability, and decision making. For instance, you might ask a visitor to try to navigate to a particular screen in an interactive and ask them to talk aloud as they do to understand their decision-making and what, if anything, is proving difficult. Doing this in real time alongside the visitor allows you to hear their gut reactions to the content or choices presented, watch their gestures and actions, and also ask follow-up questions in the moment as you watch them make choices to better understand what is going through their minds.

Observations: There are two categories of observations – naturalistic and standardized – and both can be used in formative evaluation. Naturalistic observations involve taking open-ended notes on visitors’ behaviors. Applied to formative evaluation, this might mean asking visitors to use a prototype and taking notes on their specific actions while doing so. While the notes are open-ended, they should relate to your core questions. Standardized observations record behaviors in a standard manner (e.g., a checklist). They are less common in formative evaluation but could be used to test usability. For instance, a checklist can capture whether visitors were able to navigate to particular

screens in an interactive, or complete certain tasks.

Short-answer interviews: Interviews lend depth to a formative study and are essential for answering questions relating to meaning-making. In formative evaluation, interview questions focus on the specific design concepts being tested. For instance, you might ask questions designed to assess visitors’ preferences for certain graphics, their understanding of the big idea of an exhibition, or a prototype’s ease of use. Importantly, interviews in formative evaluation are short – they should take no longer than 20 minutes. The goal is to allow visitors to tell you about their experiences using a prototype in their own words, but it is not an opportunity for deep reflection.

A combination: Usually, using a combination of these methods will best answer your questions. A/B testing, think-aloud protocols, and observations are most effective when paired with short-answer interviews. Using more than one method allows you to answer both questions of behavior and meaning-making.¹²

Evolving Approaches

No matter which strategies you use, learning through formative evaluation is exciting – the rapid, iterative process often feels like light bulbs going on or like solving a riddle. But how might we evolve and strengthen our traditional approaches to formative evaluation? This is where user experience (UX) and information architecture can offer some ideas.

¹² Ibid.



The University
of the Arts

MFA MUSEUM EXHIBITION PLANNING + DESIGN

For over a quarter century, MFA students have investigated museums and their role in society. Real-world projects combined with studio training, research, and prototyping prepare graduates to emerge as cultural advocates and problem solvers.

Learn more at: uarts.edu

UArts MEPD led
exhibit workshop
in Buenos Aires,
Argentina

EXPANDING
THE HUMAN EXPERIENCE



www.blue-telescope.com
hello@blue-telescope.com
212-675-7702



solid|light

The American Civil
War Museum
Opening 2019



We design and build destinations where people
connect with stories and each other.

solidlight-inc.com



HealyKohler Design

Museum Planning Exhibition Design Media Design Environmental Graphics

www.healykohler.com

202-774-5555

With its origins in web design, user experience (UX) encompasses all aspects of an end-user's interaction with a company, its services, and its products.¹³ Professionals working in UX have pioneered thinking on a person's holistic experience with an organization, and in particular, usability, or the ease of use of products or services. Nielsen Norman Group, a leading UX research firm, identifies five primary attributes to usability: learnability, efficiency, memorability, errors, and satisfaction.¹⁴ As exhibitions incorporate more digital experiences, we might look to UX for guidance on testing these distinct attributes more systematically.

Other lessons might come from a related area, called information architecture (IA): the practice of deciding how pieces of a whole should be arranged to best communicate to intended users, guided by principles of how people best process information.¹⁵ IA focuses on organizing, structuring, and labeling content in an effective way. Again, while rooted in the web, IA principles ultimately transcend mediums, and testing for them could enhance visitors' meaning-making within exhibitions. For instance, conducting "tree testing" – a usability technique traditionally used to evaluate the "findability" of topics on a website – for exhibitions could help museum professionals more systematically understand whether the hierarchy of information within an exhibit component is well-structured and if it makes sense to visitors.¹⁶

Further, while UX and IA employ several methods we already use in formative evaluations (e.g., Think-aloud Protocols, A/B Testing, Interviews), they also use methods rarely seen in formative evaluations for exhibitions. For instance, Desirability Testing, a card-sorting exercise used to identify first-impressions of a visual design.¹⁷ Thinking critically about how formative evaluation, UX, and IA overlap can enhance our appreciation of the many and varied possibilities for testing, and elevate our ability to evaluate and improve exhibition designs. Better design leads to more potential for meaning-making, which should always be our goal for exhibitions.

Conclusion

As you can probably tell, I am enthusiastic about the many approaches to and possibilities for using formative evaluation in exhibition development. That is because ultimately, successful formative evaluations help create powerful exhibitions by bringing visitors' thoughts, opinions, and actual experiences to the forefront of the design process. Hopefully, this article has helped demystify formative evaluation and helped you think about how you might use it to strengthen your next exhibition.

Cathy Sigmond is Research Associate at RK&A, Inc., a planning, evaluation, and research firm with offices in Alexandria, Virginia, and New York City. sigmond@rka-learnwithus.com

13 "The Definition of User Experience (UX)," *Nielsen Norman Group*, accessed December 14, 2018, <https://www.nngroup.com/articles/definition-user-experience/>.

14 "Usability 101: Introduction to Usability," *Nielsen Norman Group*, accessed September 5, 2018, www.nngroup.com/articles/usability-101-introduction-to-usability/.

15 Abby Covert, *How to Make Sense of any Mess* (Middletown: Abby Covert, 2018), 166.

16 Treejack (www.optimalworkshop.com/treejack) is one tool commonly used by IA professionals for tree testing.

17 Bella Martin and Bruce Hanington, *The Pocket Methods of Universal Design* (Beverly: Quarto Publishing Group, 2018), 29; "Using the Microsoft Desirability Toolkit to Test Visual Appeal," *Nielsen Norman Group*, accessed September 5th, 2018, www.nngroup.com/articles/microsoft-desirability-toolkit/.