2024

Conceptualizations of Data Visualization Use Beyond Efficiency in Evaluation

Sarah Douville
Claremont Graduate University

Conceptualizations of Data Visualization Use Beyond Efficiency in Evaluation

By
Sarah Douville

Claremont Graduate University
2024

Approval of the Dissertation Committee

This dissertation has been duly read, reviewed, and critiqued by the Committee listed below, which hereby approves the manuscript of Sarah Douville as fulfilling the scope and quality requirements for meriting the degree of Doctor of Philosophy in Psychology with a concentration in Evaluation and Applied Methods.

Tarek Azzam, Chair
Claremont Graduate University
Senior Visiting Fellow &
University of California Santa Barbara
Professor

Stewart Donaldson
Claremont Graduate University
Distinguished University Professor

David Fetterman
Claremont Graduate University
Adjunct Professor, Visiting Scholar, and Research Fellow

Thomas Archibald
Virginia Tech
Associate Professor

Abstract

Conceptualizations of Data Visualization Use Beyond Efficiency in Evaluation
By
Sarah Douville

Claremont Graduate University: 2024

Data visualization (data viz) is a valuable tool within evaluation for its ability to aid cognitive efficiency over text-based presentation of data (Card et. al., 1999; Evergreen, 2017, 2018; Few, 2012; Nussbaumer Knaflic, 2015; Tufte 2001). This exploratory multi-phase mixed methods research study considers purposes for using data viz in evaluation that can be achieved with that *increased efficiency* through the research question: “What conceptualizations of data viz *use* do program evaluators have beyond increased efficiency?”

In Phase I, secondary analysis of existing interview data with experts in both data viz and evaluation was used to better understand conceptualizations of data viz and their prevalence in evaluation. Support within evaluation was established for a *Utilization-Focused Evaluation Framework* (*U-FE*) (Patton & Campbell-Patton, 2022) conceptualization of data viz; an *explain* ←→ *explore* model that considers data visualization from the perspective of who is having the experience along a continuum between *explain* and *explore* (Evergreen & Metzner, 2013; Kirk, 2019), and a model on a continuum from *data* → *insight* familiar to the fields of computer science and cognitive science that converts *data* into *information*, *knowledge*, *understanding*, *sense-making*, and/or *insight* (Chen, 2009).

A model of using data viz for *stakeholder or audience engagement* and to extend evaluation *use* emerged and was further described using follow-up interviews in Phase II. This *audience engagement* model resembles the information processing model of memory (Huang et al., 2009) from an evaluation perspective. In this model, attracting and holding audience attention is intended to lead to connection (interaction) and memory (learning), which in turn leads to evaluation *use*. The model also considers evaluation specific conceptualizations of the role(s) that brand identity, data viz design principles,

artifacts, capacity building, professionalism, credibility, satisfaction, and confidence may play in audience engagement and evaluation *use*.

In Phase III, three of these conceptual frameworks (*explain* ←→ *explore*, *data* → *insight*, and *audience engagement*) were presented as brief explainer videos to a sample of 131 evaluators who are members of the American Evaluation Association to determine their familiarity with and perceived usefulness of the models. Findings suggest that there is significant conceptual overlap between the models. All three models are complementary, appropriate in evaluation, add value to the *efficiency rationale* of data viz, make sense to evaluators, and are considered useful in evaluation. Each has the potential to benefit evaluators as they consider *why* they should *use* data visualization in their work and evaluators provided many examples of using each model in their work.

This research supported that program evaluators *usually* (38.2%) or *always* (42.7%) use data viz in their evaluation work, accept the *efficiency rationale*, and are interested in other reasons for using data viz beyond *efficiency*. While data viz is a time-consuming skill, providing evaluators with conceptualizations of data viz beyond *efficiency* may make them more willing to expend the time and effort needed to apply data viz to their evaluation work. Participant interest in both the content and the medium (e.g., brief explainer videos) suggests that there is interest, need, and desire for more professional development in data visualization and associated skills. Beyond skills workshops and "how to" guides, findings suggest a desire for more learning opportunities about abstract concepts, which offers new opportunities for teaching experiences and professional development opportunities within the profession.

Overall, findings suggest that the *explain* ←→ *explore* model is a simple framework that an evaluator can use to consider the purpose of a particular visual before beginning to design and the *data* → *insight* model is a linear description of how to get the most information and insight out of a particular data viz. The *audience engagement* model is a holistic approach to thinking through the relationships in the evaluation to support evaluation use. While there is no clear hierarchy of models suggested in this study, comments supported that the *audience engagement* model is the most specific to evaluation – to the extent that it might not even be data viz specific.

Table of Contents

List of Tables

List of Figures

Why Use Data Visuals in Program Evaluation?

Program evaluators serve an important translational role in making often complex knowledge accessible to stakeholders (Valéry, 2007). By organizing data visually and then leveraging human biology and cognitive processes to assist in discovering and understanding findings, evaluators and stakeholders can benefit from data visualization (data viz) in a multitude of ways. Unsurprisingly, data viz is a popular topic within evaluation with a Topical Interest Group of the American Evaluation Association (AEA) devoted to it, a large number of data viz specific presentations at AEA's annual conference, and a number of evaluators publishing books, websites, blogs, and workshops on the topic such as Ann Emery (https://depictdatastudio.com) and Stephanie Evergreen (https://stephanieevergreen.com).

Most data viz sources advocate including data viz for the efficiency gains realized by moving slower text reading processes into faster pattern-recognition processes, or for the aesthetic or attractive qualities of data viz (Evergreen, 2017, 2018; Few, 2012; Nussbaumer Knaflic, 2015; Tufte 2001). These reasons are applicable to evaluation and many of the data viz choices evaluators make can be summarized within one of those two contexts. For instance, increased efficiency could be articulated as "make the data easier or faster to understand," "share a large amount of data," "accurately compare," or "draw extra attention." Visual appeal for the audience could be articulated as "break up a text report," "save space over presenting as a table," or "support the stakeholder's branding design." On their own, these two rationales are likely sufficient reasons to include data visualization in evaluation work.

However, even if efficiency and visual appeal are sufficient, the evaluator may have other reasons for including viz that do not so easily fit into these categories. And, while the first two areas have been written about extensively in evaluation, a number of other potential reasons to include data viz have not been thoroughly explored within evaluation. For instance, that same evaluator may use a data viz to persuade the audience, make the data more memorable, encourage evaluation use, satisfy a client to encourage future contracts, allow interactive features that allow the audience to engage deeply and generate their own questions, etc.

The ostensibly simple question: “Why use data viz in program evaluation?”, reveals its complexity when a program evaluator consciously considers the many potential reasons to use data viz. Further, these intended uses must be carefully balanced because, while a single data viz may fulfil several purposes, it is unlikely that a single data viz serves *all* purposes as some design choices naturally conflict with other purposes (Hegarty, 2011). For example, an embellished graph may be more memorable or persuasive (Batch, 2018; Rykaczewska, 2021), but it may not be as efficient to read; a fully interactive visual may draw the audience away from the story being told and into asking questions not relevant to the evaluation questions (Lysy, 2013); and a simple or efficient visual may not be culturally relevant for the audience (Azzam et al., in press).

While other fields may be content to use data viz simply based on the rationale that it aids in efficiency, program evaluation has an emphasis on the concept of *use* and a tradition of being intentional and reflective in practice. This dissertation will explore potential conceptual frameworks (e.g., mental models) that program evaluators can use to consider the many reasons to use data viz that go beyond simple efficiency.

Review of Literature and Presentation of Key Conceptual Models

There are a number of reasons to use data viz and a number of ways to conceptualize those reasons for including data viz. Although there are no generally accepted typologies of data viz *use* (Börner et al., 2019; Brehmer and Munzner, 2013; Cavaller, 2021; Pfitzner, 2001; Zhu, 2007), and certainly none within evaluation, this literature review will describe the overarching reason for using data viz (to aid efficiency), offer a definition of data viz in evaluation, describe a framework of evaluation *use* and populate that framework with examples applicable to data viz, and then present two complementary conceptual frameworks for considering the purpose of data viz (*explain* ←→ *explore* and *data* → *insight*) within evaluation.

Data Visualization Aids Efficiency and Specific Tasks

Data visualization draws from multiple fields including cognitive science, computer science, statistics, user experience, communication sciences, and psychology (Aparicio and Costa, 2014 in Cavaller, 2021). These fields have different research traditions, different language to describe and measure data viz, and different conceptual models of visualization including a number of taxonomies and suggested typologies. Many of these sources purport benefits of data viz and offer reasons for including data viz consistent with the over-arching rationale accepted in this paper: to aid efficiency.

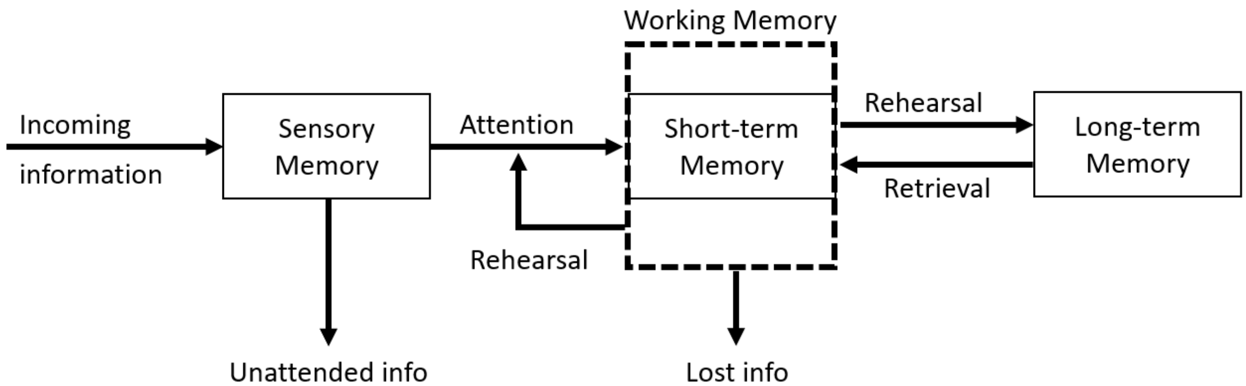
To Aid Efficiency or Reduce Cognitive Load

Ware (2021) describes the human visual system as a "flexible pattern finder coupled with an adaptive decision-making cognitive mechanism" (p. 2). When an explicit reason for using data viz is given, a variation of improving efficiency consistent with Ware's description of the human visual system is usually offered. The primary arguments are based on harnessing strengths of human biology such as presenting data in ways that our powerful, pre-attentive visual system can find patterns easily. This frequently offered reason for using data viz is often framed as increasing processing efficiency for the user and is often combined with the highly inter-related concept of cognitive load.

First described within educational psychology in the 1980s, cognitive load theory refers to the functions of and relationship between working memory and long-term memory (Sweller et al., 2019). Figure 1 is a clear description of this relationship as presented visually by Huang and colleagues (2009).

Figure 1

The Information Processing Model of Memory



Note. This is a recreation of the visual presented by Huang et al. (2009, p. 3)

Cognitive load theory differentiates between *intrinsic* cognitive load (the inherent difficulty of the information being presented) and *extraneous* cognitive load (difficulty ascribed to the way in which the information is presented). Intrinsic cognitive load also considers the individual characteristics of the person receiving the information such as domain knowledge, cognitive processing ability, personality traits including locus of control, and previous experiences (Moore, 2017). Such an understanding of the audience is highly relevant within data viz design in evaluation (Azzam et al., in press). Extraneous cognitive load focuses on the format and presentation of the material (Huang et al., 2009) and reducing this type of cognitive load is a common focus of data viz designers.

Reducing cognitive load is expected to be achieved by converting intensive cognitive processes such as reading text into faster perceptual processes such as pattern recognition (Pfitzner, 2001) to support transferring knowledge from working memory into long-term memory (Sweller et al., 2019; Ware, 2021). Most data viz sources recommend simplifying materials to avoid overwhelming the recipient's working memory with statements such as "the main difference between effective and ineffective data displays is their inability to communicate the evaluator's key message in a clear and straightforward way such that it does not overload a viewer's working memory capacity" (Evergreen & Metzger, 2013, p. 6).

Terminology associated with improving efficiency and reducing cognitive load include quicker or faster (Evergreen & Metzner, 2013) or easier to perceive (Ware, 2012 in Mason and Azzam, 2019). Many other provided rationales such as supporting decision-making (Moore, 2017), amplifying cognition (Card et al., 1999), augmenting cognition (Hegarty, 2011), or assist reasoning (Tufte, 2016) can be interpreted as also relying on improved efficiency or reducing cognitive load.

Proponents often promote data viz by pointing out that vision is human beings' strongest sense and visual processing is among our strongest cognitive strengths (Few, 2012; Ware, 2021). Specifically, the human brain is excellent at recognizing patterns and data viz supports this strength by organizing and "pre-chunking" data for pattern recognition (Evergreen, 2011; Freedman & Shah, 2002; Jones et al., 2019; Shah & Carpenter, 1995). In addition to an intuitive acceptance of visual processing as an innate human strength (e.g., consider the old adage that a picture is worth a thousand words), a wealth of research in cognitive science supports that we are vision-based animals and that vision is our most efficient way of making sense of the world (Few, 2012; Hegarty, 2011; Ware, 2021).

Task Specific: Obtain a Specific Data Point or Piece of Information

A great number of sources conceptualize the goal of data viz as supporting the completion of a specific task such as determining a specific data point or locating a specific piece of information (Conati et al., 2014; Pfitzner, 2001; Zhu, 2007). Again, this framing is usually underpinned by the assumption that data viz aids in efficiency, speed, and reduces cognitive load as described above. In fields other than evaluation, a great deal of data viz work is built around defining tasks and measuring task efficiency (Zhu, 2007).

An early organization of task types was Shneiderman's (1996) "type by task taxonomy" of information visualization tasks consisting of gaining an *overview* of the entire collection, being able to *zoom* in on items of interest or *filter* out uninteresting items, selecting *details-on-demand* of an item or group, viewing *relationships* among items, keeping a *history* of actions, or *extracting* sub-collections. He indicated that these "seven tasks are at a high level of abstraction" and encouraged others to refine the tasks (p. 3423). A number of authors then attempted to differentiate between high-level and low-level tasks such as Conati and colleagues (2014) who described low-level task types such as retrieve value, find extremum, sort, and compute derived value; while describing high-level tasks as "subjective open-ended decision making activities" (p. 5).

These tasks are often described as verbs – the actions that one is taking with the data – such as "sort" (e.g., evaluator sorts program sites by highest posttest scores), "compare" (e.g., an average line allows stakeholders to compare their program site to others), or "present" (e.g., the evaluator uses the data viz to tell a story). While these tasks are meant to be abstract and domain independent, there is little agreement between fields and classification systems as to the "appropriate granularity" (Brehmer and Munzner, 2013, p. 1) of a task such that one article may refer to extracting a value as a task (Shneiderman, 1996) and another may refer to making a decision (Conati et al., 2014) as a task. While those intent on measuring task performance do need to concern themselves with defining and measuring these disparate conceptualizations of task, one of the conceptual frameworks described later, the *data* → *insight* continuum, may ameliorate this need for evaluators. The *data* → *insight* continuum progresses from relatively concrete concepts such as data and information to more abstract concepts such as insight with an emphasis on moving from *data* to *insight*. I suggest that this conceptualization is more relevant to evaluators when conceptualizing the various reasons for using data viz than focusing on defining and measuring tasks.

However, a problem central to this research still remains. Conceptualizing tasks as verbs implies that the action one is taking is the same as *why* they are doing it. Action such as "sort" or "compare" can be very descriptive, but not very useful in understanding a high-level purpose. In critiquing one of the more robust descriptions – Roth's (2012) taxonomy which distinguishes between *goals*, *objective*, *operators*, and *operands* within visualization tasks – Brehmer and Munzner (2013) note that the highest level of the taxonomy, “goal” uses the verbs *procure*, *predict*, and *prescribe*, and "does not provide us with any higher-level context or motivation for *why* the user is procuring" (p. 2376). Brehmer and Munzner (2013) go on to stress the need to differentiate between the means ("*how* the task is performed, and *what* are the task’s inputs and outputs") and the ends ("*why* the task is performed") of a data viz task (p. 1). These ends are the *why* this dissertation is concerned with. In organizing *why* a task is performed, Brehmer and Munzner (2013) suggest many of the components found within the data viz *use* frameworks that I will discuss. For instance, within their high-level *why* category of *consume*, Brehmer and Munzner (2013) list *present* (*explain*) and *discover* (*explore*), which are the anchors of the explain ←→ explore model described later. They also provide a category for *enjoy* which is reminiscent of art, aesthetics, or attracting users, which will be a large component of the *audience engagement* model that emerges from this research and is described later. This dissertation is concerned with *why* a task is performed, as raised by Brehmer and Munzner (2013), but from an evaluator perspective.

Efficiency is Sufficient, but Intellectually Limiting

Harnessing human biology to improve efficiency (Few, 2012; Ware, 2021) is an entirely appropriate underlying reason for using data viz. It holds strong face validity and is supported by both research and simple biology (Hegarty, 2011; Ware, 2021). This dissertation accepts this underlying reason but contends that it should not be the sole consideration in the multifaceted field of evaluation.

A major drawback of focusing only on the efficiency of data viz is that we may place ourselves in a research and design paradigm emphasizing efficiency as if that were the *only* consideration, thus blinding ourselves to other potential benefits of data viz. Zhu (2007) points out the circular nature of data viz efficiency literature in which researchers strive for additional methods of measuring efficiency because they assume efficiency is what they should measure; without questioning what else should be measured. In accepting the superiority of data viz that maximize efficiency, we may also blind ourselves to unintended consequences of those viz choices. For instance, the bar char is used frequently in evaluation and often lauded for its efficiency (Evergreen, 2017; Evergreen, 2018; Few, 2012), yet recent research by Holder and Xiong (2022) suggests that a downside of the ubiquitous bar chart is that it may encourage social bias and related stereotyping by hiding variability within the data being presented.

An over-emphasis on efficiency may also lead us toward less creative data viz and fewer experimental data viz in the design world. Kirk (2019), as well as Tufte in his more recent works, have emphasized the artistic value of data viz and expressed concern over embracing a dreary world overly reliant on bar charts. Further, an over-emphasis on speed and accuracy can limit our understanding of other *uses* for data viz. As Freedman and Shah (2002) state, “accurate and fast fact-retrieval is less critical than a deep understanding of relationships and trends” (p. 2). Considering the drawbacks to focusing only on the efficiency of data viz, this research seeks to focus on conceptualizations of data viz *use* that go beyond efficiency and could be applicable within evaluation.

Definition of Data Visualization in Evaluation

This dissertation uses a definition of data visualization adapted from Kosara (2007) that should be familiar within evaluation and has been used in other evaluation specific data visualization research (Jones et al., 2019; Wanzer et al., 2020). This definition is articulated by Azzam and colleagues (2013) as: “Data visualization is a process that (a) is based on qualitative or quantitative data and (b) results in an image that is representative of the raw data, which is (c) readable by viewers and supports exploration, examination, and communication of the data” (p. 9).

There is considerable overlap between data viz and graphic design. Considering that evaluators are communicators and knowledge brokers (Jones et al., 2019) and both data viz and graphic design can support communication efforts (Evergreen, 2001), the fields of data viz and graphic design complement each other and should both be considered when creating evaluation deliverables. This research will restrict the definition of data viz to the visual display of *data* and will not address reasons to focus on formatting and graphic design in evaluation. For a thorough discussion on formatting and graphic design in evaluation, read Evergreen (2011, 2017, 2018).

Inclusion of qualitative data in this definition may cause confusion for some readers in differentiating graphic design elements from data visualization. In these cases, consider the context, intent, and the underlying data. For instance, a stock photo of smiling children on the cover of a nutrition program evaluation report is likely a graphic design decision. Photos within an evaluation report of children displaying the meals they were served by the program are likely data visualizations meant to describe the food served. Photos of meals could also be considered qualitative data that is then analyzed during the evaluation. Further, one might question whether some visuals of qualitative data are “an image that is representative of the raw data” or an image of raw data post data analysis. Considering the intent of the visual and that many visuals of qualitative data are further along the *data* → *insight* continuum, as discussed later, can help resolve this conceptual conflict.

Some visuals remain potentially difficult to categorize. For example, information graphics (info viz) and icons are sometimes based on underlying data and should be considered data viz (e.g., using a smiley face to denote positive sentiment from qualitative feedback) or may simply be stylized design features of the report (e.g., using the program’s logo as a custom bullet point). Again, context, intent, and underlying data should be used to differentiate between data viz and graphic design for the purposes of this discussion.

Visualizations of Quantitative Data

The bulk of data viz effort, discussion, and tools are devoted to quantitative data, which typically means data that is conceived as or presented as numeric. For many sources, the term data viz is synonymous with quantitative data viz and no space is devoted to qualitative data. This is consistent with the emphasis on data viz within heavily quantitative fields such as computer science and statistics, and the role of tracking data points which are commonly expressed numerically such as quantities, stock prices, etc.

While quantitative and qualitative research traditions are philosophically and methodologically distinct (Plano Clark & Sanders, 2015), the distinction between quantitative and qualitative data is much more subtle (Sandelowski et al., 2009). Even the term "raw data" used in the definition of data viz above can cause confusion since a number without context is not data (Few, 2019) such that even "raw" data must include an entity and a relationship to be considered data (Bertin, 1977 as cited in Ware, 2021). These entities are typically qualitative categories such as temperature, age, or fruit; and additional attributes may be added such as the color of the fruit (Ware, 2021). Therefore, to be understandable, quantitative data have qualitative contextual features.

Some visuals lend themselves equally well to quantitative or qualitative data (e.g., a call-out box or boxed display could contain words, images, or numbers) and some, such as joint displays, intentionally combine qualitative and quantitative data in one display. However, many visuals are more common or appropriate to either qualitative or quantitative data. For instance, a scatterplot is a common representation of quantitative data but is not common (and likely inappropriate) for qualitative data (Verdinelli & Scagnoli, 2013) while a word cloud is commonly used with qualitative data.

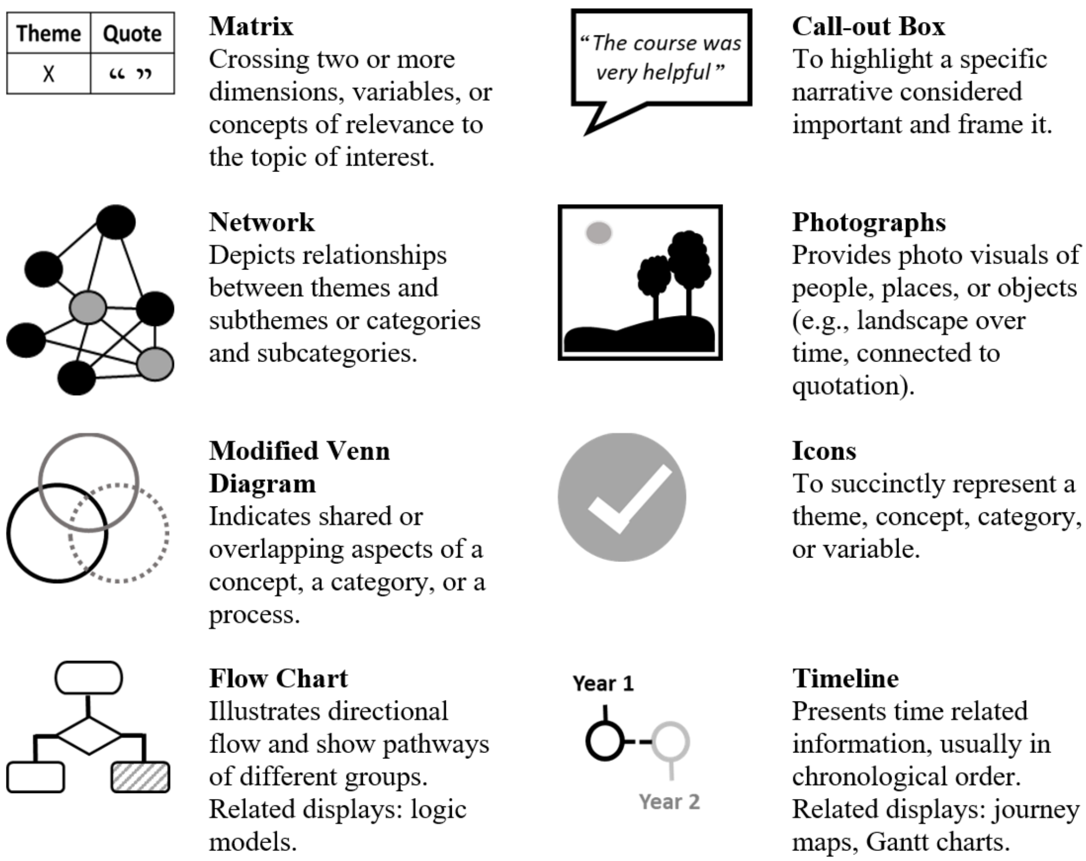
Visualizations of Qualitative Data

Qualitative data are very often displayed using the same graphs used for quantitative data displays. This is typically done by first *quantitizing* or "assigning numerical (nominal or ordinal) values to data conceived as not numerical" (Sandelowski et al., 2009, p.2) to the qualitative data and then treating it the same as quantitative data by displaying as a bar chart, pie chart, etc. The pro-offered reasons for doing so follow the same arguments offered for visualizing quantitative data; increasing efficiency and harnessing the human ability for pattern recognition. One ubiquitous qualitative data specific visual, the word cloud, *quantitizes* the frequency of particular words and denotes frequency with text size (Jones et al., 2019). Word trees and phrase nets (Henderson et al., 2013) are two other visual displays that express *quantitized* qualitative data using text size but these visuals also suggest relationships to other words by including information on proximity to other words.

There are a number of visualizations commonly used with qualitative data that do not simply rely on *quantitizing* data. Examples of these data visualizations from Azzam and colleagues (in press) are provided in Figure 2. For instance, timelines can visually display key events or information in a logical order (Verdinelli & Scagnoli, 2013) and iconic displays (Hegarty, 2011) can represent complex physical objects such as a diagram of a machine or a map of the world. Network diagrams, modified Venn diagrams, and flow charts can be used to illustrate *relationships* between concepts, groups, or workflows (Azzam et al., in press; Verdinelli & Scagnoli, 2013). These types of visualizations are sometimes referred to as information visualization instead of data visualization (Kirk, 2019). This distinction is interesting because it implies that the output (the image) is from “information”, a process further along the *data* → *insight* continuum than simple “data”.

Figure 2

Example Visualizations of Qualitative Data from Azzam and Colleagues (in Press)



Visualization of qualitative data is less commonly discussed than quantitative data (Henderson & Segal, 2013) but is prominent in this paper due to (a) the importance of qualitative data in evaluation and (b) this project’s heavy reliance on visual conceptual models, which can be considered a visualization of qualitative data. One potential complicating factor in working with qualitative data is that significant pre-processing must be done to move it from *data* to *information* or *knowledge* before visualizing it and a tremendous amount of this preprocessing is currently being done by human beings (Ware, 2021). Another factor likely contributing to the dominance of quantitative data in visualization is that it is often perceived as more credible or analytical than qualitative data (Slone, 2009 in Henderson et al., 2013).

Program Logic Models as Visualizations of Qualitative Data

One of the most recognizable data visualizations within evaluation is the program logic model (Jones et al., 2019). Logic models are visual representations of the relationships between program components and program outcomes (Chen, 2015; Donaldson, 2007; Weiss, 1998). They are typically comprised of at least two inter-related theories, the *program process theory* explicating the inputs and activities associated with the implementation of the program, and the *program impact theory* explicating the expected outcomes of the program (Donaldson, 2007). Additional theoretical frameworks may be presented with or integrated into the logic model to explain the underlying social and behavioral science theories expected to contribute to the general program theory (Donaldson, 2007).

As with other visualizations of qualitative data, logic models can support both ends of the *explain* ←→ *explore* continuum and can be used actively in communication with stakeholders and also in participatory processes as a method of co-developing a shared understanding of the program being evaluated. The W. K. Kellogg Foundation (2004) clearly articulates this use of logic models:

> In general, logic modeling can greatly enhance the participatory role and usefulness of evaluation as a management and learning tool. Developing and using logic models is an important step in building community capacity and strengthening community voice. The ability to identify outcomes and anticipate ways to measure them provides all program participants with a clear map of the road ahead. . . Because it is particularly amenable to visual depictions, program logic modeling can be a strong tool in communicating with diverse audiences – those who have varying world views and different levels of experience with program development and evaluation.

Recommended Practices in Data Visualization

Related to the growing emphasis on data visualization for communication, decision-making, and analysis, there is a growing field (and related industry) of data viz training materials, software tools, and consultants. There is also a growing body of research literature within evaluation about how to create good viz or manipulations of viz to test designs (Jones et. al., 2019; Mason & Azzam, 2019; Wanzer et al., 2020). Most of these sources focus on harnessing biological human strengths related to visual perception (Few, 2012; Tufte, 2001; Ware, 2021) and leveraging human cognitive processing abilities to emphasize pattern recognition and reduce effort on working memory (Evergreen, 2011). This paper will not outline nor debate those suggested best practices. The goal of this research is to move beyond efficiency to consider the purposes – within the context of program evaluation – we can achieve with that efficiency.

The Importance of *Use* in Evaluation

While traditional research may be conducted with the intent to add to the body of human knowledge, without a specific context or intended application, evaluation is rooted in context and conducted with an intended application (Alkin & King, 2016). That evaluation findings will be *used* in decision-making is a fundamental expectation of the evaluation field (King & Alkin, 2019). Evaluation literature often discusses the frustration that emerged in the 1970s when prominent evaluators realized and began discussing the dearth of specific examples of their evaluations being *used* in decision-making (King & Alkin, 2019; Patton & Campbell-Patton, 2021; Weiss, 1998).

Conceptualizations of what constitutes *use* were developed and expanded through this discussion. For instance, an early emphasis on linking a specific decision to a specific finding in order to demonstrate *instrumental use* expanded to allow for *enlightenment use* whereby the findings are integrated into the recipient's understanding of the topic and affect distal decision-making in more subtle ways (Kirkhart, 2000; Weiss, 1998).

Utility as a key focus within evaluation was formalized in the 1980's when the Joint Committee on Evaluation Standards highlighted *utility*, *feasibility*, *propriety*, and *accuracy* as the four required features of all evaluations (Stufflebeam, 1980 in Patton & Campbell-Patton, 2021). The Joint Committee's four required features with an emphasis on *utility* as central to evaluation is also incorporated into the popular Centers for Disease Control and Prevention framework for program evaluation in public health (CDC Framework, 1999). The importance of *use* in evaluation has elevated it to one of the most researched and discussed topics in evaluation (Patton & Campbell-Patton, 2021) and continues to be a primary focus in evaluation.

Definitions of Use in Evaluation

The language of *use* in evaluation ranges from the term *utilization* which is focused on "specific primary intended users for specific, intended uses" (Patton & Campbell-Patton, 2021. p. 5) to the word *influence* with its much broader scope of meaning including intangible and indirect affects (Kirkhart, 2000). As one might expect of a multidisciplinary field (Valéry, 2007), the field of evaluation is rife with disagreements over terminology. This is likely due in part to evaluation's use of common language to describe technical terms and exacerbated by the many different meanings and changing definitions asserted by particular authors over time. To reduce this complexity, this paper relies on the *Utilization-Focused Evaluation* (*U-FE*) framing of *use* as described so thoroughly by Patton, precisely because it has been so thoroughly described and is so well known within the evaluation field. For clarity, the word *use* is italicized when used as a concept.

Three Complementary Frameworks for Considering Data Visualization *Use*

While there is no unifying and generally accepted data viz framework (Brehmer and Munzner, 2013; Cavaller, 2021; Pfitzner, 2001; Zhu, 2007) a number of conceptual frameworks have been described or alluded to by various authors across multiple fields. Presented below are three conceptualizations applied to program evaluation. The *Utilization-Focused Evaluation* model (*U-FE*) applies the *U-FE Framework* to data visualization. The *explain* ←→ *explore* model (*EE*) is a bi-directional continuum representing the experience one is expected to have with the data between *explain* and *explore*. The *data* → *insight* model (DI) is a directional continuum of *data* into *insight* that focuses on the process of converting data into a useable product. Importantly, there is a great deal of overlap in these frameworks and all rely heavily on the underlying assumption of increased efficiency through reduced cognitive load.

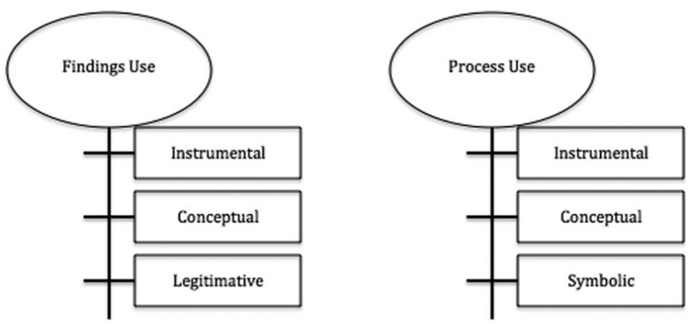
Model UFE: Data Visualization Use Applied to the U-FE Framework of Use

Although no models specifically applying the *U-FE Framework* to data visualization were found in evaluation literature, it is relatively straightforward to describe this model. The first significant definitional division within *U-FE* to consider is the distinction between *findings use* and *process use*. *Findings use* is the original conceptualization of *use* and refers to the decisions, actions, or changes made by an individual, program, or organization based on the evaluation findings (Patton, 2008; Patton & Campbell-Patton, 2021). In data viz, *findings use* would likely relate to data viz within an evaluation report or other presentation of findings. As analysis has already been conducted, it is most likely that *data viz findings use* would be closer to the *explain* end of the *explain* ←→ *explore* continuum: data viz created by the evaluator to *explain* findings to stakeholders. *Process use* is a newer conceptualization – closely tied to evaluation capacity building – that considers the changes (i.e., cognitive, attitudinal, behavioral, skill) that happen within the individual, program, or organization based on the evaluation process (Patton & Campbell-Patton, 2021). *Data viz process use* is less discussed within evaluation with the AEA Topical Interest Group focused on data viz clearly associated with findings and reporting as evidenced by its full name: Data Visualization & Reporting TIG, but this is an area of rich possibilities. Either the evaluator, or stakeholders, or both may change their understanding, attitudes, or skills through creation of, or interaction with, data viz during the evaluation.

Findings use and *process use* are separate types of *use*; neither is a sub-set of the other (Alkin & King, 2016; Patton, 2008; Patton & Campbell-Patton, 2021). Within each, we can consider separate categories of *use*. Johnson, et al. (2009) provide clear explanation of the most commonly presented categories: *instrumental use* (direct action based on evaluation knowledge), *conceptual or enlightenment use* (no direct action taken, but understanding has changed), or *symbolic use* (mere existence of evaluation is used to persuade or convince). The strong distinction drawn between *findings use* and *process use* by Alkin and King (2016) can be observed in Figure 3, a visual representation they adapted from Alkin & Taut (2002), in which they present these categories as very separate ovals without a connecting line.

Figure 3

Visual Presentation of Findings Use and Process Use Depicted by Alkin and King (2016)



Within *findings use*, *instrumental use* refers to instances where users can document the specific way(s) they used the social science knowledge for decision-making or problem-solving purposes (Rich, 1977 in Alkin & King, 2016). This category has been referred to as "allocative, direct use" (Braskamp, 1982 in Alkin & King, 2017). An example of *instrumental data viz findings use* would be a funder viewing a graph of program site pretest-posttest performance gains and choosing to defund the lowest performing program. Obvious confounds exist in this and other examples of data viz *use* as it is highly unlikely that the funder made the decision based solely on the visual, without reading the written report.

Conceptual use of findings occurs when knowledge gained from evaluation findings influences thinking which informs [later] decisions (Rich, 1977). Weiss (1977) referred to this use as *enlightenment use* and the terms are often used interchangeably. *Conceptual data viz findings use* could take the form of program staff learning from the findings of one program and later integrating those findings into the design of future programs. For instance, staff could learn that parents bring their children to the reading program in large part because they value the healthy snacks provided. One of those staff is hired by a math program and goes on to prominently feature healthy snacks in that new program.

A third category within *findings use* was described by Knorr (1977) as *symbolic use*. Also called *legitimative use* (Alkin & King, 2016), this category of *use* refers to relying on the findings of the evaluation to justify a decision already made. While using evaluation findings to support program decisions may be completely appropriate, intent probably matters most in symbolic use (Mark & Henry, 2004) and evaluators would consider many examples of use within this category as misuse (Alkin & King, 2017; Douville, 2019). In many cases, data can be provided in text or within a simple table (Azzam et al., 2013; Batch, 2018) and studies have shown that visualizations have little impact on simple tasks (Zhu, 2007), so it is not always necessary to add data viz to aid in efficiency. An example of *symbolic data viz findings use* could be an evaluator adding data viz to the report as a design decision, simply to break up the text and make the report more attractive and appealing to readers.

Within *process use*, *instrumental use* often refers to direct changes made within the program or individual participants in the evaluation as a result of participating in the evaluation. For instance, a participant might learn how to write better survey questions or gain new skills in statistics while working with the evaluation team. Or, a program might change the schedule of its workshops in order to allow time between sessions for surveys to be collected. An example of *instrumental data viz process use* would be an evaluator using conditional formatting to quickly identify missing data during analysis.

Conceptual use or *enlightenment use* of process refers to internal changes, often long-term, within the individual or organization as a result of participating in the evaluation process. Patton and Campbell-Patton (2022) equate these changes to "the difference between learning how to learn versus learning substantive knowledge about something" (p. 32). This includes changes in thinking and understanding such as learning to think critically, learning to employ the logic of evaluation in decision-making, or having new insight into one's implicit bias. Especially in evaluation approaches that heavily include stakeholders in evaluation processes such as *U-FE*, participatory evaluation, developmental evaluation, etc., c*onceptual use* may be a primary and intentional aspect of the evaluation. *Instrumental use* and c*onceptual use* are closely tied to Evaluation Capacity Building (ECB) and may be framed as such (Patton, 2008). An example of *conceptual or enlightenment data viz process use* would be increasing visualcy (also referred to as graphical literacy) of stakeholders through data viz explanations and exercises during the evaluation process (Azzam et. al., in press; Börner et al., 2019; Zhu, 2007). A visualization of qualitative data specific to evaluation, the logic model, could also provide *conceptual data viz process use* by making underlying assumptions explicit and providing stakeholders with a different understanding of their program than they had before engaging in the process of creating a logic model.

Symbolic process use is using the fact that an evaluation has been commissioned – rather than any findings or data from the evaluation itself – for some end, for instance, to make (or delay) a decision. As with *symbolic findings use*, these *symbolic process use* decisions might be completely appropriate, but they could also be misuse (Alkin & King, 2017; Douville, 2019). *Symbolic data viz process use* could be as straightforward as adding data viz because people expect it or could be the program making claims based on the mere presence of data viz such as claiming to be "data driven" simply because they have a data dashboard, regardless of its role in their decision-making.

Findings use and *process use* are distinct, but the demarcation between *instrumental use*, *conceptual use*, and *symbolic use* is blurry (Pelz, 1978, as cited in Alkin & King, 2016). Proper categorization will often be impossible because the internal processes, as with many social sciences processes, cannot be directly observed. Also, in the case of data viz, one viz may serve multiple purposes and be used differently at different times by different users. For the purposes of this paper, therefore, Figure 4 depicts *findings use* and *process use* as separate types represented by different colors and divided by a thick line. A dashed line denotes the permeability between *instrumental*, *conceptual*, and *symbolic use*.

Figure 4

Model UFE: Examples of Data Viz Use Within Findings Use and Process Use Categories

	Findings use	**Process use**
Instrumental *Allocative, direct use*	Instances where users can document the specific way(s) they used the social science knowledge for decision-making or problem-solving purposes. *Funder closes weakest site based on a graph of program site pretest-posttest performance gains.*	Direct changes made within the program or individual participants in the evaluation as a result of participating in the evaluation. *Evaluator uses conditional formatting to quickly identify missing data during analysis.*
Conceptual *Enlightenment use*	When knowledge gained from evaluation findings influences thinking and informs [later] decisions. *Chart shows dramatic increase in repeat attendance of library reading program after healthy snacks are added. The next program the library adds, to raise math scores, prominently features healthy snacks.*	Internal changes, often long-term, within the individual or organization as a result of participating in the evaluation process. *Stakeholder visualcy is increased through data viz explanation exercises during the evaluation.*
Symbolic *Legitimative use*	Relying on the findings of the evaluation to justify a decision already made. *Evaluator adds data viz to the report as a design decision, simply to break up the text and make the report more attractive and appealing to readers.*	Using the mere existence of an evaluation to justify decisions or make claims. *Program proclaims that it is "data driven" and "backed by data" because it has created a dashboard display.*

Note. The examples have been invented for illustration purposes and are not drawn from literature.

Model EE: Data Visualization Experience Along Continuum of Explain ←→ Explore

A potentially useful framework for considering data visualization *use* is along a continuum with *explain* and *explore* as anchors, Figure 5. Kirk (2019) describes this continuum as "judging the experience offered by your visualization" (p. 82) and also offers a mid-continuum category of *exhibitory*. Kirk (2019) uses this broad *exhibitory* category to describe a wide variety of reasons to create a visual such as, an elaborate data viz created upon stakeholder request when a basic chart of data would have been appropriate, visuals that appear explanatory but require text to support interpretation, or a viz that appears to be explanatory but requires the viewer to "finish the task of understanding" (p. 86).

Figure 5

Model EE: Bi-directional Continuum of Explain to Explore

In short, the ends of the continuum are easy to conceptualize but the murky area in the middle bears reflecting on. For instance, the first example provided above by Kirk (2019) would require context but could fall into a category of *symbolic use* in an evaluation *use* context (i.e., creating an elaborate viz because a stakeholder requests it). The second example, a visual that appears explanatory but requires text to support interpretation, also requires context but could be provided to make a specific detail from the text memorable or could be intended to attract or engage the audience. The third example that requires the viewer to finish the task of understanding is relying on designing the visual such that insight occurs and can be considered as leveraging *insight* for audience engagement. This three-part continuum supports Kirk's discussion and is appropriate there, but this paper will exclude the *exhibitory* category. This rich "catchall" category includes many less obvious reasons for *use* that should be considered within evaluator specific conceptualizations of *use*.

Therefore, the simpler model of *explain* ←→ *explore* will be used in this discussion. Another benefit of using this model is that it is familiar within evaluation. Evergreen and Metzner began their abstract in a *New Directions for Evaluation* (2013) issue devoted to data visualization with "data visualization is often used in two main ways—as a tool to aid analysis or as a tool for communication" (p. 1).

Explain. The furthest end of the *explain* anchor of the continuum exemplifies the adage that a picture is worth a thousand words. Brehmer and Munzner (2013) use the word *present* to describe *explain*: "the use of visualization for the succinct communication of information, for telling a story with data, guiding an audience through a series of cognitive operations" (p. 2378).

Explain data viz are commonly discussed in the context of evaluation reports (Douville et al., in press), which is consistent with the report's goal to explain findings to an audience. In evaluation reporting, the data viz is being used as a communication tool – sharing a story determined by the evaluator, possibly with the participation of other stakeholders. And while there may be untapped opportunity for evaluators to more fully engage audiences with an *explore* data viz, the evaluator would also be giving up control over the data presentation and missing an opportunity to act as a data storytelling guide (Lysy, 2013). While not all *explain* data viz is static, they often are in evaluation (Douville et al., in press).

Explore. At the other end of the continuum, *explore* suggests that the viewer (often referred to as a user in this context) is able to use the data viz to conduct or generate their own analysis, answer multiple questions with the data, and generate new questions (Azzam et al., 2013; Isenberg, 2013; Smith, 2013). *Explore* is often associated with interactivity, but this is not a necessary requirement. Additional conceptualizations of use are often embedded within a description such as the argument in favor of providing exploratory interactive graphs to turn the audience from "a passive consumer into an active participant" (Weissgerber et al., 2017, p. 20592 cited in Rykaczewska, 2021).

Explore is sometimes offered as an overall reason for data viz with statements such as "visualization is concerned with exploring data and information," and overlaps with many other scientific approaches to knowledge generation such as data mining (Chen et al., 2009, p. 12). Very closely tied to *sense-making* and *insight*, Brehmer and Munzner (2013) use the word *discover* to describe using data viz for scientific investigation through generation and verification of hypotheses.

A great deal of data exploration does not use data viz, relying on statistics instead. Batch (2018) found that “visualization is primarily seen as a communication tool among professional analysts and that few of our participants ever use visual representations of their data in the middle of an analysis process” (p. 279). Batch’s sample consisted of experienced data analysts who preferred statistics or data tables to data viz and, when interviewed, indicated that data viz was not needed, took too long, or was not as detailed as numeric data. However, some data visualization practitioners have advocated for more emphasis on data viz to assist with the analysis of data including the statistician Tukey who created several data viz for analysis including the stem and leaf plot (Evergreen, 2011). Batch (2018) also argues that we should strive to have more analysts use data viz in their analysis, not just as an explanatory tool.

The Importance of Qualitative Data. Verdinelli and Scagnoli (2013) provide a review of visualization techniques that feature qualitative data used in research and emphasize the role of creating visual representations of emerging theories as “an intrinsic and essential step in theory building” (p. 2). Data visuals are a method of articulating our mental models – the way we perceive reality – and presenting that model to others so that they can critique, correct, and build upon them. Visualizing qualitative data can support both ends of the *explain* ←→ *explore* continuum. The process of creating the visual helps the theorist to *explore* and refine their understanding of the qualitative data, while also creating an artifact that can help to *explain* their mental model to others.

Many commonly used evaluation methods such as observation, interviews, focus groups, and open-ended survey responses generate qualitative data. These data may be analyzed and/or displayed using visualizations. Visualization of this qualitative data may also play a process role in evaluation such as graphic recording (also known as graphic facilitation, visual recording, doodling, sketch noting, or viz thinking) as discussed by Dean-Coffey (2013), data placemats (Pankaj & Emery, 2016), storyboarding

(McAlindon et al., 2019), or mind mapping (Dean-Coffey, 2013). The process role of data viz has the potential to be significant within this research as we consider “who is having the experience” within the *explain* ←→ *explore* continuum.

Who is Having the Experience? Deeply intertwined with the *explain* ←→ *explore* continuum is the point of view of the person having the experience with the data viz. *Explain* ←→ *explore* refers to the experience the user is having and is independent from the underlying data. For instance, an evaluator may use data viz to *explore* a particular data set (e.g., create a histogram to inspect frequency of all categories and inspect for outliers), but the data viz produced and shared with stakeholders may be used as a communication tool to *explain* the findings (e.g., a stacked bar chart displaying only the two categories of interest to the evaluation question with all other data points and outliers grouped and grayed out as “others”).

Considering the benefits of data viz in supporting pattern recognition during analysis, there are at least two interesting potential topics of investigation within program evaluation related to *explore* data viz. First, like other analysts, evaluators may be under-utilizing exploratory data viz and missing opportunities to utilize the *data* → *insight* continuum most efficiently. Conventional wisdom estimates that up to 80% of data analysis time is spent on data cleaning prior to analysis (Dasu & Johnson, 2003), but this is not necessarily wasted time. Batch (2018) and others have suggested that *sensemaking* is happening during this process such that by the time the data are cleaned, a story is already emerging. Given evaluators’ emphasis on storytelling, it is possible that evaluators *are* conducting exploratory data visualization during data cleaning (e.g., generating scatterplots or histograms during statistical analysis, or color-coding variables during data cleaning) and simply not conceptualizing this specifically as an *explore* data viz activity.

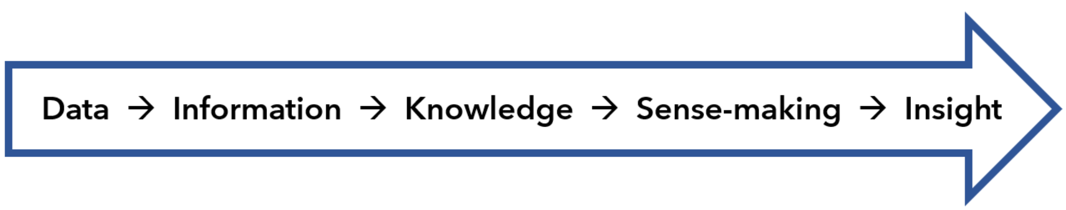
Second, the benefits of the *explore* data viz process may be unequally shared by stakeholders. In research conducted by Douville and colleagues (in press), interviews with experts in data viz suggest that evaluators are more likely to create static data viz for stakeholders that *explain* findings than to engage stakeholders in *exploring* data viz. As with the potential missed opportunity of evaluators rarely describing using data viz for their own data exploration, there is potential missed opportunity for data viz use if evaluators are not engaging their audience in exploring data viz.

Model DI: Continuum of Data → Information → Knowledge → Sense-making → Insight

Closely tied to the rationale of improving efficiency is an argument that data viz helps turn *data* into *information* or *knowledge* (by reducing complexity, supporting pattern recognition, etc.) and that doing so allows even greater human cognitive processing of that *information* or *knowledge* into the often-purported ultimate goal of data visualization: *sense-making* or *insight*. While the continuum itself is not commonly discussed within eval, there is obvious overlap with the often mentioned role of evaluation to “tell a story with data.” These states can be conceptualized along a continuum from *data* to *insight* as shown in Figure 6.

Figure 6

Model DI: Continuum of Data → Insight



As with nearly all the terms explored here, definitional differences exist across fields and between authors for each of these words. Rather than belabor the definitions and fixate on the exact dividing point between each term, this continuum is offered in the spirit of the Oxford Dictionary (2022) definition of continuum as a “continuous sequence in which adjacent elements are not perceptibly different from each other, although the extremes are quite distinct.” Chen and colleagues (2009) indicate that a general "consensus exists that *data* isn’t *information* and that *information* isn’t *knowledge*" (p. 12). Accepting this continuum as a mental model or typology, rather than as a strict taxonomy allows us to accept ambiguity in the definitions while still agreeing, for instance, that *knowledge* is further along the continuum than *information*, and *insight* is even further from *data* than *knowledge* is from *data*.

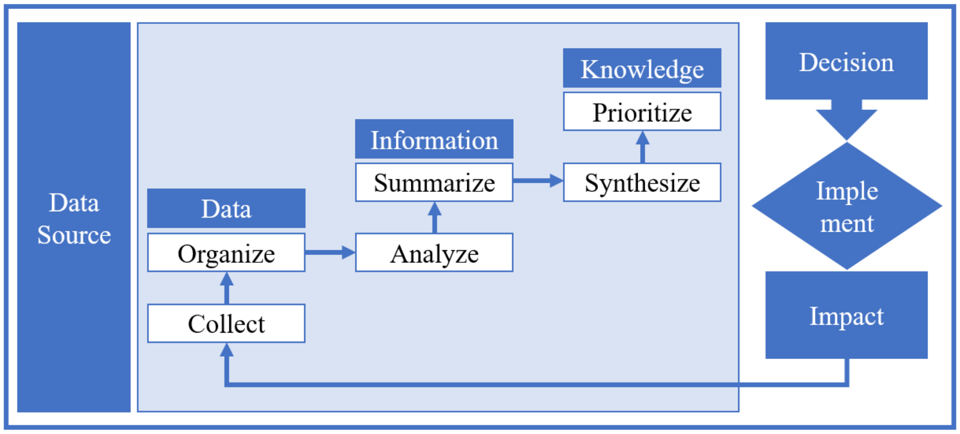
While various sources suggest different transitions along this route (e.g., *information* to *knowledge*; *knowledge* to *insight*; or *data* directly to *insight*) there is always implied directionality (e.g., there is no discussion of turning insight into information nor knowledge into data) and a general consensus that there are definitional distinctions between the terms (Chen et al., 2009).

Move From Data to Information to Knowledge. An individual data point is rarely useful. Typically, processing is required to add relationships and context to convert data into information. For instance, a participant's pre-test score is simply a data point, but comparing that score to all other participant scores provides information about the participant's performance. Similarly, a computational process such as statistical analysis can convert multiple data points into a data point plus information (e.g., participant score and average score) which is also information rather than data. Chen and colleagues (2009) describe an iterative loop of processing from *data* → *information* → *knowledge* and indicate that a key tenant of knowledge is that human cognitive processes, typically described as reasoning, is necessary to convert *information* → *knowledge*.

Data viz is often promoted as a method to display the patterns and relationships that turn data into information. However, creating *information* or *knowledge* is not typically a desired end goal in evaluation, *use* of that information or knowledge is expected. The framework for data driven decision-making described by Mandinach et al. (2006) discussed by Moore (2017) and presented as Figure 7 may be more applicable to program evaluation. This model depicts the first portion of the continuum, showing *data* → *information* → *knowledge* to support making a decision. This decision action could result in *instrumental use* if used directly, or *conceptual use* if the knowledge informs future decisions. In this context, *conceptual use* is likely more aligned with *sense-making* as described next.

Figure 7

Framework for Data Driven Decision-Making



Note. This figure has been recreated from the figure presented in Mandinach et al., 2006 as cited by Moore, 2017.

To Support Sense-Making. Sense-making is the act of "interpreting" data, information, or knowledge and "inventing" a story, framework, or mental model that fits the data (Yi, 2008) and this term is used frequently in evaluation. The term sense-making sometimes describes the entire process from initial data inspection to presenting to an audience (Lee, 2016). Sense-making is also used to describe the *process* by which the analyst makes a "conscious effort" (Lee, 2016) to apply "their knowledge and understanding to interpret meaning from data" (Batch, 2018).

Outside of program evaluation, this process is very often described as a sensemaking loop using terminology similar to descriptions of how people build mental frameworks. For instance, in the data/frame theory of sense-making articulated by Klein and colleagues (2006), the user already has a pre-existing framework into which they fit the data they are inspecting and then either confirms and elaborates the frame by placing agreeing evidence into it or reframes by discarding and rebuilding the framework when disconfirming evidence is found. This model is conceptualized as a loop of sensemaking activities that may be cycled through multiple times (Lee et al., 2014). Pirolli and Card

(2005) describe this process as *information* → *schema* → *insight* → *product* and also suggest that it is iterative.

The iterative nature of this process fits well within the *data* → *insight* continuum in that *data* can be converted into *information* and then more *data* or more *information* can be added to convert that *information* into *knowledge*. Integral to all these explanations is the role of a human being in the sense-making loop, which parallels the evaluator's role in facilitating shared understanding, which is also a stated program evaluator competency (American Evaluation Association, 2018).

To Encourage Insight. Creating *insight* is often proposed as the reason for visualization (North, 2006). The problem is, once again, that the definition of *insight* differs by academic field and sometimes even by the context of the study. Fortunately, the two most prominent definitions, from the cognitive science and the computer science schools of thought, do not contradict one another and both can be applied to the program evaluation context.

The word *insight* as it is used in the cognitive science community refers to the spontaneous shift from "a state of not knowing how to solve a problem to a state of knowing how to solve it" in an "aha" or "eureka" moment that does not follow a traditional problem-solving process (Chang et al., 2009, p. 14). In this tradition, *insight* is associated with increased neural activity in specific areas of the brain. Functional Magnetic Resonance Imaging (fMRI) technology has documented that traditional problem-solving relies on increased – and sustained – neural activity in brain areas associated with working memory, executive processes, and long-term memory; *insight* relies on these areas *and* is accompanied by a sudden burst of neural activity associated with the dopaminergic midbrain's reward process (Chang et al., 2009; Tik et al., 2018). This definition of *insight* implies that the process is something that takes place within a human brain and literally changes the person.

Data viz literature is heavily influenced by computer science literature and in this tradition, *insight* is more often conceptualized as an understanding that can be uncovered. The *insight* is an artifact that exists and which the user finds through analysis. For example, in their research interviewing business analysists, Kandogan and colleagues (2014) state that their goal was to "understand how data insights are created and delivered to [business decision-makers]" (p. 3), implying that once uncovered it is like any other knowledge that can be taught to another person. Chang and colleagues (2009) articulate this difference between research traditions and refer to "spontaneous" insights and "knowledge-building" insights as distinct (p. 17) while arguing that visualization of data can be used to support *insight* in both traditions.

General Research Objective

This research study seeks to better understand *why* program evaluators use data visualization in their work. The overarching research question is: What conceptualizations of data viz *use* do program evaluators have beyond increased efficiency?

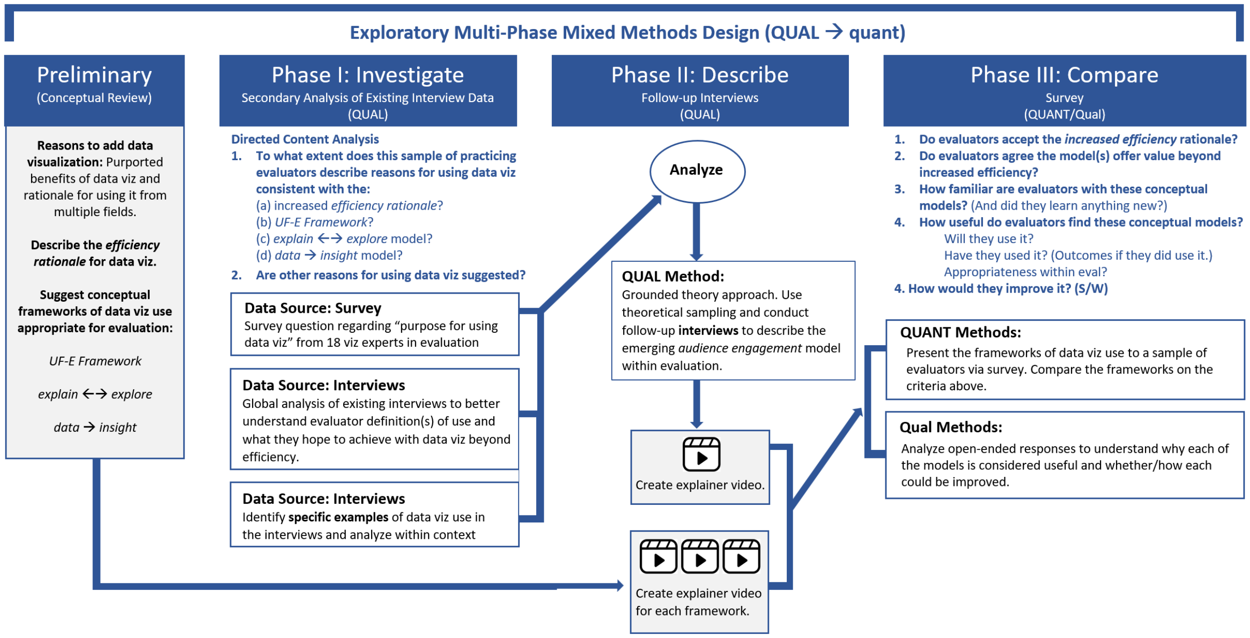
If conceptualizations of data viz *use* beyond increased efficiency are apparent within evaluation, those conceptualizations will be presented to a sample of evaluators to understand their familiarity with the models and their perceived usefulness of the models to evaluation.

General Method

This research project uses an exploratory multi-phase mixed methods design (Creswell & Poth, 2018) in three phases, see Figure 8. Qualitative data is prioritized in Phases I and II. Quantitative data is prioritized in Phase III. Phase I investigates evaluator conceptualizations of data viz *use*; Phase II was added to the study to describe a newly uncovered conceptualizations of data viz *use*; and Phase III compares these conceptualizations of data viz *use* with a sample of practicing evaluators to better understand which, if any, of the conceptual models are considered helpful within evaluation.

Figure 8

Visual Depiction of Proposed Study Design



Phase I

Three potential models of conceptualizing reasons for using data viz within evaluation were suggested following a review of literature within and outside evaluation. The first model, based on the *Utilization-Focused Evaluation Framework* (Patton, 2008; Patton & Campbell-Patton, 2022), is not specifically discussed in evaluation literature within the context of data viz, but should be familiar to evaluators given the prominence of the *U-FE Framework* within evaluation (King & Alkin, 2019). The second conceptual framework of *explain* ←→ *explore* is discussed within evaluation literature (Evergreen & Metzner, 2013) and is predicted to be familiar to evaluators. The third conceptual framework of *data* → *insight*, familiar within computer science and cognitive psychology literature (Chang et al., 2009), is not directly discussed within evaluation literature but concepts consistent with the definitions of *sense-making* and *insight* from the *data* → *insight* model can be found within evaluation literature and practice.

This phase uses a secondary analysis of existing data to consider whether there is support for the three models under consideration and to search for any additional models. The existing data for analysis is from a 2021 study (Douville et al., in press) of surveys and interviews with experienced program evaluators exploring the *process* that data viz creators engage in when developing viz.

Phase I Research Question

Phase I seeks to address these specific research questions:

1) To what extent does this sample of practicing evaluators describe reasons for using data visualization consistent with the:
    a. commonly accepted overarching rationale of increased *efficiency*?
    b. *Utilization-Focused Evaluation Framework*?
    c. *explain* ←→ *explore* model?
    d. *data* → *insight* model?
2) Are other (previously unidentified) reasons for using data visualization suggested by this sample of practicing evaluators?

Phase I Method

Douville and colleagues (in press) conducted a 2021 study of data viz experts in evaluation and other fields to explore the challenges and limitations evaluators experience when creating data viz. Two datasets from that study, (1) a short survey and (2) eighteen interviews, are analyzed in the present study.

Secondary Analysis of Existing Survey and Interview Data

This study is a secondary analysis conducted on existing data, rather than a secondary data analysis. Data are traditionally considered secondary data when they are made publicly available by the original researchers for external analysis (Pienta et al., 2011). The data in this study have not been made publicly available. They are being used by one member of the original research team with the consent of the original research team (Appendix A) and are being used to answer a separate research question.

As suggested by Cheng and Phillips (2014), the term *secondary analysis of existing data* is used in this study in place of the traditional and potentially confusing term *secondary data analysis* to clearly indicate that these data are being used to answer different research questions than they were originally collected to answer. One portion of the original dataset, the survey question asking: *What is your purpose for using data visualization?* was not analyzed by the original research team. The second dataset used here, eighteen of the qualitative interviews, was analyzed by the original research team into themes related to evaluation challenges, but conceptualizations of data viz *use* were not investigated by that team. This study provides an opportunity to investigate new phenomena and answer original research questions that have not been addressed before with this dataset.

Phase I Participants

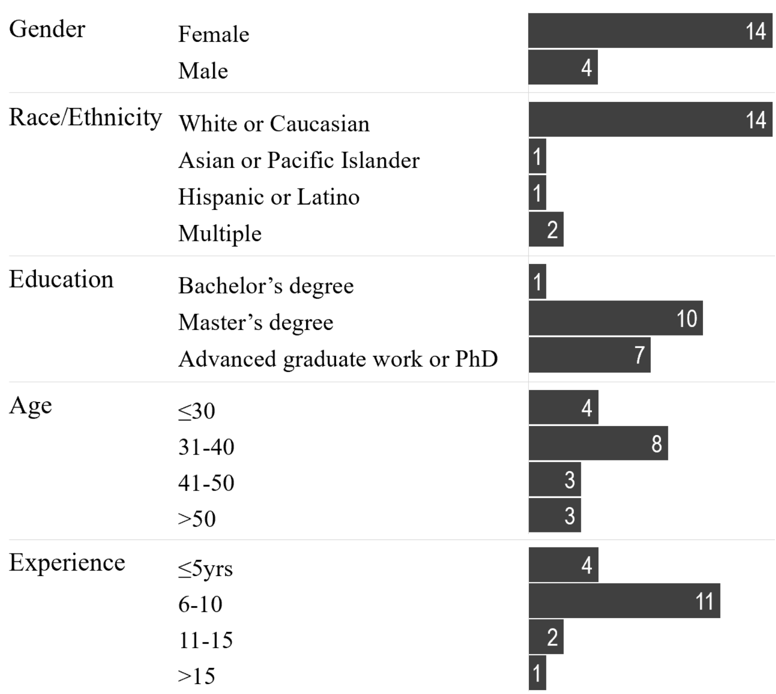
Phase I participants are (N = 18) program evaluators who are experts in data visualization. A purposive reputational sample (Bamberger & Mabry, 2020) of data viz experts were invited to participate in the 2021 study. Participants were identified based on reputation within the data viz field including: holding a position associated with expertise such as being a data viz topical interest group leader for a professional association; having published books, articles, blogs, or podcasts on the topic; having taught courses on the topic; having conducted presentations or workshops at professional conferences on the topic; or through snowball sampling (Creswell & Poth, 2018) whereby the participant is recommended as an expert by another identified expert. Participants were recruited primarily via email, with several participants approached via Twitter or LinkedIn. A total of 70 individuals were contacted at least twice and 26 of these individuals completed the interviews (37% response rate). The participants from other fields were removed from the dataset and only the data from the eighteen participants identified as *program evaluators* were considered and analyzed for this phase.

The eighteen participants included were categorized as evaluators for having been a member of the American Evaluation Association (AEA) or because they specifically self-identified as an evaluator during the interview. Evaluator participants ranged in age from 28 to 56 years (*M* = 38.8; *SD* = 8.8), were primarily white, female, well educated, and from North America (see Figure 9). Self-reported years of data viz experience ranged from 3 to 30 years (*M* = 9.5; *SD* = 6.0).

To maintain anonymity, participants are assigned a random double letter identifier followed by their years of evaluation experience such that II-6 is an evaluator with six years of experience as of 2021.

Figure 9

Phase I Participant Demographics



Phase I Materials

There were two data sources for this secondary data analysis: 1) Participant responses to an open-ended survey item and 2) participant interviews. All assertions regarding confidentiality, data storage handling, data storage, etc. from the prior Institutional Review Board review [CGU Protocol #3882] remain in place and are honored in this secondary analysis.

Open-ended Data Viz Use Survey Item

Participants completed a brief survey (Appendix B). In addition to reporting demographic information (e.g., age, race, gender, years of data viz experience, etc.), participants completed an open-ended, short answer question that asked, *What is your purpose for using data visualization?* Responses to this question ranged from 5 words to 49 words (M = 19.4; SD = 11.4).

Data Viz Use Interview

The 2021 interview protocol (Appendix C) was designed to better understand evaluator *experiences* creating data visualizations. Participants were asked to outline their process for creating data visualizations and to pay particular attention to the challenges they experienced. All interviews were conducted during Spring 2021 via video conferencing to comply with COVID-19 pandemic precautions. Those interviews ranged from 41 minutes to 88 minutes in length (M = 55.0; SD = 12.5).

Phase I Data Analysis

Directed content analysis was used to code both responses to the open-ended survey item and interviews. Directed content analysis (Hsieh & Shannon, 2005) is a structured and deductive approach to exploring pre-defined categories within existing data sources. While directed content analysis is deductive, it also allows new themes to emerge, if relevant. Assarroudi and colleagues (2018) suggest that using "predetermined code" and "category" interchangeably can be confusing (p. 46), so I will use "category" or "model" to refer to the existing conceptual models of interest (*increased efficiency*, *U-FE Framework*, *explain* ←→ *explore*, and *data* → *insight*) and "code" to refer to components of the *a priori* models. When used as a verb, the word "code" or "coding" refers to the process of assigning segments of text to either a "category" or "code." In this way, a segment of text may be coded directly to a category such as the *data* → *insight* model or to a code within that model.

Hsieh and Shannon (2005) recommend beginning a directed content analysis by either highlighting potentially relevant parts of the text and then returning to code them or starting the coding process before initial highlighting. I chose the latter approach because I was already very familiar with the data. Therefore, I reviewed the data carefully and began coding, using the models as categories and adding additional codes as needed. In addition to looking for instances of the *a priori* models within the data, I paid attention to any reasons for using data viz that did not clearly fit within the *a priori* models. New reasons that emerged were turned into codes and applied to subsequent data sources.

While the overarching method of this is a directed content analysis, the data were organized into three different data sets: (1) responses to a short survey question, (2a) specific examples of data viz *use* extracted from the eighteen qualitative interviews, and (2b) eighteen qualitative interviews in their entirety. Analysis of each of the three data sets was handled slightly differently to better understand if there are differences in the prevalence of models in these different data sets. Each route of analysis prioritized finding support for the models in the dataset, erring intentionally on the side of identifying potential cases. Support for the models is defined in Table 1.

Table 1

Definition of Prevalence (Extent to Which the Model is "Apparent") in Each Data Set

	Survey Responses (N = 18)	Use Cases (N = 29)	Interviews (N = 18)
prevalence is determined by…	mention of word(s) or synonyms closely associated with the model of interest	mention of word(s) or concepts associated with the model of interest	the model can be discerned in at least half the interviews and… similarity to terminology as used in the literature
model is considered…			
very apparent	≥75% of participants reference it	≥75% of *use* cases reference it	model is easily identified and the terminology used is very similar to literature
apparent	between ≥50 and 75% of participants reference it	between ≥50 and 75% of *use* cases reference it	model or major elements of the model are clearly present but the terminology used by participants is different from literature
slightly apparent	fewer than 50% of participants reference it	fewer than 50% of *use* cases reference it	elements of the model are discernable – possibly with generous interpretation – but the descriptions and terminology are not similar to literature
not apparent	no participants reference it	no *use* cases reference it	elements of the model cannot be identified

Data Viz Use Survey Item Analysis

Survey responses to the question *What is your purpose for using data visualization?* were reviewed and examined multiple times, considering specific word choices and possible alternative meanings, to determine whether the evaluators suggested a purpose for using data visualization consistent with the *a priori* models. In this analysis, mention of word(s) or synonyms closely associated with the model of interest is sufficient to indicate that the model is referenced. Special attention is given to any responses falling outside one of these *a priori* categories as such responses could suggest other conceptualizations of data viz *use* appropriate to evaluation. Consistent with a project on data viz, visually arranging the categories, color-coding, and formatting were used to support analysis. Analysis was conducted in Microsoft Excel. The model is considered *very apparent* if ≥75% of participants reference it, *apparent* if between ≥50 to 75% of participants reference it, *slightly apparent* if fewer than 50% of participants reference it, and *not apparent* if no participants reference it.

Specific Examples of Data Viz Use

The interview data were organized two different ways for analysis. First, specific *use* examples provided by the participants were extracted from the interviews and each example of *use* was read multiple times, considering specific word choices and possible alternative meanings, to determine whether the evaluators suggested a purpose for using data visualization consistent with the *a priori* models. In keeping with the spirit of the prior 2021 study, this process was intended to elicit specific examples with their associated context.

Coding and identification of specific use examples was conducted in MAXQDA 2022 (VERBI Software, 2022) and analysis was done in Microsoft Excel. As some participants contributed more than one example of data viz *use* for analysis, prevalence of the model in this dataset is determined by the percent of specific *use* cases that can be interpreted as referencing the model, not by the number of participants referencing the model. The model is considered *very apparent* if ≥75% of *use* cases reference it, *apparent* if between ≥50 to 75% of *use* cases reference it, *slightly apparent* if fewer than 50% of *use* cases reference it, and *not apparent* if no *use* cases reference it.

Global Coding Interview Approach

Second, each interview was analyzed to theme comments related to data viz *use* whether related to a specific example or not. The specific *a priori* categories discussed so far as models were applied to the interview data with each model considered a parent category. A broader global coding approach was undertaken to achieve a more holistic understanding of the data while seeking a balance between identifying instances of *a priori* categories and identifying new themes. Each interview was read multiple times to determine whether the evaluators suggested a purpose for using data visualization consistent the *a priori* models at any point in their discussion. Coding was done in MAXQDA 2022 (VERBI Software, 2022). As additional codes arose, their intended meaning was analyzed across subjects to better understand if the codes represent new models or if they can reasonably be sub-themed within an existing model. Some of these codes were straightforward (e.g., the *explain* ←→ *explore* model included separate codes for *explain* and *explore*; the *U-FE framework* included codes for *instrumental*, *conceptual*, and *symbolic*); and the *data* → *insight* model included codes for *learning*, *understanding*, *knowledge*, *insight*, *storytelling*, and *decision-making*.) However, there is great overlap between the models and some codes are associated with multiple models. In such cases, context, the participant's other related statements, and researcher familiarity with the models guided placement of a coded segment within a parent category. For instance, "to aid in *decision-making*" is a commonly expressed reason for using data viz but is not a stand-alone model in this study. When participants gave that reason, I used context to place that example within either the *U-FE Framework,* the *data* → *insight* model, or both. As with other analysis approaches in Phase I, I erred on the side of assigning a coded segment to a model to support the presence of the model rather than excluding it. Examples of coded segments that could represent multiple models, the model(s) I placed them in, and my rationale can be found in Table 2.

Table 2

Examples of Coded Segments that Potentially Apply to More than One Model with Rationale

Participant Quote	Model(s) Represented	Appearance	Rationale
…we're trying to talk about the quantitative data while mixing it with the qualitative insights that we had from people, and we felt this was a really good depiction of what are these low engagement people like, so the top half is more the quantitative outcomes and the bottom half is more the qualitative from the focus group [AA-10]	*data* → *insight*	*slightly apparent*	discussing complex integration of data to move from data toward something more complex
	explain ←→ *explore*	*apparent*	"trying to talk about" and "depiction" suggest *explain*
…really generate knowledge and transfer knowledge that's actionable and people are actually gonna use to learn and improve [CC-30]	*data* → *insight*	*apparent*	generating knowledge
	U-FE	*apparent*	acting on knowledge
…showing the gap between intent and actual…when I can prove myself and let them know how evaluators can be useful for them…I can bring data to you, and look at how beautiful this data is, that's gonna help your decision-making [GG-10]	*audience engagement*	*apparent*	attract attention, promote credibility
	data → *insight*	*slightly apparent*	showing an insight
	U-FE	*slightly apparent*	decision-making

As all models can be discerned in at least half of the interviews, albeit with generous interpretation and substantial overlap, purely quantitized counts of references to the model are not appropriate in this analysis. Here the definition of *apparent* relies heavily on my perception of the effort required to extract the model from the data and the extent to which interviewees are describing the model (or components of the model) in alignment with the terminology as it is used in literature. Here, I applied a broad synthesis of my readings of the data and denote the model *very apparent* if the model is easily identified and the terminology used is very similar to literature, *apparent* if the model or major elements of the model are clearly present but the terminology used by participants is different from literature, and *slightly apparent* if elements of the model are discernable – possibly with generous interpretation – but the descriptions and terminology are not similar to literature. Table 3 in findings provides further details and an overview on how apparent each model is within the data.

Phase I: Findings

There were two data sets analyzed, (1) responses to a short survey question asking: *What is your purpose for using data visualization?* and (2) eighteen qualitative interviews with evaluators who were experts in data viz. This second data set was analyzed using two different approaches, (2a) identification and categorization of specific examples of data viz *use* and (2b) holistic analysis of the interviews. Table 3 displays how apparent each model is within the data by route of data analysis. Findings from the three analyses are described here separately, but integrated in the discussion and conclusion.

Table 3

Model Prevalence in Survey Responses, Use Cases, and Interviews

	Survey (*N* = 18)		Use Cases (*N* = 29)		Interviews (*N* = 18)	
Model	*n*	%	*n*	%	%	Prevalence considering similarity to literature
Efficiency Rationale	11	61.11	11	37.93	>50	*very apparent*: frequently used terminology such as easy, fast, quick, etc.
U-FE Framework	5	27.78	26	89.66	>50	*very apparent*: frequently used terminology directly from UFE literature
Explain ←→ Explore	14	77.78	15	51.72	>50	*very apparent*: language suggesting *explain* to stakeholders dominated with less frequent mention of *explore*
Data → Insight	7	38.89	12	41.38	>50	*slightly apparent*: limited mention of *insight* but strong emphasis on knowledge, learning, and remembering
Other	13	72.22				
Audience Engagement	11	61.11	28	96.55	>50	*very apparent*: language suggesting *audience engagement* and the sub-themes identified were very apparent, but had not been previously reviewed in literature

Survey Findings

The survey question *What is your purpose for using data visualization?* was answered by all eighteen participants, see Table 4.

Table 4

Definitions of Use (Purpose) Provided via Survey

Evaluator	What is your purpose for using data visualization? (Distilled)	Efficiency Rationale	U-FE Framework	Explain ←→ Explore	Data → Insight	Other
AA-10	...To convey complex processes or results, in a way that **explains** the underlying construct or finding(s) that is more **efficient than text**. They are also an avenue for creating consensus on a topic that facilitates further discussions.	Yes		Explain		avenue for creating consensus on a topic that facilitates further discussions
OO-10	to **communicate information clearly and efficiently** using charts, graphs, plots, images and infographics.	Yes		Explain (communicate)		
YY-5	Broadly, the purpose of using data visualization is to **effectively and efficiently communicate data** that meets the varying needs of the target audience(s).	Yes		Explain (communicate)		meets the varying needs of the target audience(s)
II-6	To me, the goal in visualization is to **convey information** (data) to an audience in an **clear, effective, succinct, compelling, and easy-to-understand** way	Yes		Explain (convey)		compelling
JJ-3	Make reports/products **more interesting and visually appealing**; makes data **more digestible** to broader audiences	more digestible		Explain		make more interesting and visually appealing
LL-4	To **better show** trends, summaries, and the overall message of information than what a table or paragraph could do.	to show better than a graph		Explain (better show)	overall message of information	
XX-8	To make **data from lengthy tables and dataset more accessible**, with an emphasis on tailoring visualizations around specific audiences.	more accessible than tables		Explain (more accessible)		accessible with emphasis on tailoring
ZZ-10	To make information **attractive and easily consumed**.	easily consumed		Explain (easily consumed)		attractive
BB-15	To **better communicate complex** information.	communicate complex information		Explain (communicate)	complex information	

Evaluator	What is your purpose for using data visualization? (Distilled)	Efficiency Rationale	U-FE Framework	Explain ←→ Explore	Data → Insight	Other
CC-30	Data visualization is a **communication** tool that leverages what we know about human vision to **complement verbal and written communication**.	leverage human vision		Explain (communication tool)		communication
FF-10	to **engage** stakeholders and clients; to **summarize lots of data in an easy-to-digest display**	summarize lots of data, easy to digest		Explain (display)		to engage
EE-4	...If we **want people to care about using data and understand what the data says**, then we have to figure out ways to present that information to them that is **easily understood, enticing, and eye-catching**	no	care about using data and understand	Explain (present)	understand what the data says	get people to "care" by being "enticing, and eye-catching"
MM-7	To **elucidate communication** such as to **maximize an audience's understanding** of the data's significance	no	maximize understanding	Explain (communicate)	maximize understanding	
TT-8	To explore data or explain data, often for reporting purposes for clients.	no		To explore data or explain data		reporting purposes for clients
KK-10	It allows data to come to life, and if done well, enables people to **better understand and make decisions** with data.	no	make better decisions		make better decisions, better understanding	allows data to come to life
VV-15	**Increased understanding** of content / findings / analysis to **inform decisions** that can **improve participant outcomes**.	no	inform decisions, improve outcomes		increased understanding, inform decisions	
GG-10	Helping my clients to **make sense of their data**, using it for more **informed decision-making**, keeping up with my peers ;-)	no	more informed decision-making		to make sense; more informed decision-making	keeping up with my peers ;-)
WW-6	To make data more **approachable, engaging, and relevant** to varied audiences	no				approachable, engaging, and relevant

Note. Two responses were abbreviated, as indicated by ellipses, but no words were changed. Bold was added to align with the conceptual models suggested. Dark green fill is used to indicate that the word efficiency is used in the definition; light green indicates that similar words were used. Orange was added to draw attention to mention of audience.

Increased Efficiency Rationale

The rationale that data viz increases efficiency is *apparent* in the survey data. Over 60% (*n* = 11) of the sample described *increased efficiency* as a rationale for using data viz. Four of eighteen specifically mentioned the word *efficiency* and a further seven used language that suggested *efficiency* such as "show better than [table or text]" or "easy to digest." Of the six that did not mention *efficiency*, four of these were among the minority of people who gave answers that did not align with the *explain* ←→ *explore* continuum either.

U-FE Framework

Reasons for using data visualization consistent with the *U-FE framework* were only *slightly apparent* in the survey data. There were no direct references to the *Utilization-Focused Evaluation framework* within the survey responses. Erring on the side of supporting model prevalence, several responses that reference "decision-making" are coded as aligning to the *U-FE framework* as these could be *instrumental use* within the *U-FE framework*. However, these decision-making comments could also be interpreted as closely aligned with the high-level data task of "decision-making," as described within the data task frameworks discussed by Brehmer and Munzner (2013), Shneiderman (1996) and others or the knowledge decision-making framework described by Mandinach and colleagues (2006). There were five comments that mentioned versions of "better understanding" that could be related to *conceptual use*, but since no comments mentioned distal use such as "learning that can be applied to later decisions," from the *U-FE framework* there is not a clear fit with this aspect of the *U-FE framework*, either. There simply isn't enough context in the short survey responses to support a finding greater than *slightly apparent*.

Explain ←→ Explore Model

The *explain* ←→ *explore model* is *very apparent* in the survey data. Consistent with the literature review demonstrating that the concept of *explain* ←→ *explore* is likely familiar to evaluators, fourteen (77.78%) of the evaluators in this sample described *use* in a way that aligns with the *explain* ←→ *explore* continuum. The majority interviewed here aligned toward the *explain* end of the continuum. One participant identified the continuum directly and said the purpose of data viz is "to explore data or explain data" [TT-8]. Twelve (66.67%) of the evaluators used either the word "explain" in their purpose or similar language such as "communicate", "convey", or "show." Of the five evaluators who did not clearly align with the *explain* ←→ *explore* continuum, two described *use* around engaging stakeholders and three described *use* along the *data* → *insight* continuum as "helping my clients to make sense of their data" [GG-10], "increased understanding…informed decisions" [VV-15], or "better understand and make decisions" [KK-10].

Data* → *Insight Model

The *data* → *insight model* is *apparent* in the survey data. No participants used the specific term *insight* to describe their reason for using data viz, but one referenced sense-making [GG-10], and five specifically mentioned *understanding*, which is along the *data* → *insight* continuum. Three of the evaluators specifically mentioned *decision-making* as a purpose for data viz which, in addition to being a high-level data task, can also be embedded within the decision-making framework (Mandinach et al., 2006, cited by Moore, 2017) associated with the *data* → *insight* model. Admittedly, the emphasis on finding support for *a priori* models in the data results in identifying the *data* → *insight* continuum as *apparent*. If, for instance, the decision-making framework presented by Mandinach and colleagues had been an *a priori* code, it would likely have been *apparent* as well.

Emerging Model Emphasizing Audience

A second study question asked whether there are *other reasons for using data visualization suggested by this sample of practicing evaluators?* A number of comments suggested that a theme of using data viz to attract and engage an audience is *apparent* in the survey data. Eleven of the 18 (61%) respondents specifically mentioned audience, client, or stakeholder in their responses. Relevant terms used were coded orange in Table 4 to highlight this prevalence. The fact that survey responses generally aligned with the *explain* end of the *explain* ←→ *explore* continuum also suggests an emphasis on *audience* even when audience is not specifically mentioned. Emphasizing *explain* suggests that evaluators were conceptualizing data viz as a way to *explain* data to an audience rather than considering the evaluators themselves as the *users* of the data viz. Even the one respondent that specifically mentioned *explore* as a reason then added "often for reporting purposes for clients" [TT-8].

Triangulating Evidence with an Artificial Intelligence Summary

While not a formal method used in this study, the survey responses lend themselves well to summary by generative artificial intelligence software. The survey responses to this question were entered into the Open.ai ChatGPT model, which produced a summary of that dataset that clearly reflects efficiency (e.g., clear, succinct) and also clearly emphasizes the importance of audience engagement. The model was instructed to consider only the dataset provided in its response. Interestingly, the summary does not use the words *efficient*, *explain*, nor *explore* even though the quotes submitted use these terms heavily.

> The goal of data visualization is to effectively communicate information using charts, graphs, images, and infographics in a clear, succinct, and visually appealing way. It helps to show trends, summaries, and the overall message of information that cannot be conveyed through tables or paragraphs. The purpose is to engage stakeholders and clients, make data more digestible to broader audiences, and to better communicate complex information. By using data visualization, we can make data more approachable, engaging, and relevant to varied audiences and inform decision-making (ChatGPT 3, personal communication, March 5, 2023).

Interview Findings

Interviews were analyzed in two different ways. First, specific examples of *use* were extracted and considered within their specific context to better understand what participants consider examples of data viz *use*. Next, the interviews were reread in their entirety and coded for general references to conceptualizations of data viz use. In both analyses, the expected evaluator emphasis on utilization was *very apparent*, even though it was only *slightly apparent* in the shorter survey responses.

Specific Examples of Data Viz Use

All except one of the evaluators interviewed supplied at least one specific example of their data viz being *used* in practice and elaborated on what *use* meant to them. Appendix D outlines those twenty-nine examples of *use* along with evidence of *use* and reasons why they thought their data viz contributed to that *use* – if a reason was given. Most evaluators (*n* = 10) offered one specific example; three offered two examples; three offered three examples; and one evaluator offered four separate examples of data viz use. The one evaluator who did not provide a specific example of use during the interview responded, “It's hard for me to answer that question, since I'm not the user of the evaluation findings ... I almost feel like it's the clients who need to be answering it…We know that our clients really like what we put together generally, but we don't know, always like exactly how it's being used” [JJ-3]. The interpretation of use suggested by this person is consistent with a *U-FE* focus on *findings use*, but potentially complicates the conceptualizations of data viz *use* provided by other participants, as will be discussed later.

Increased Efficiency. The rationale that data viz *increases efficiency* is only *slightly apparent* when examining the specific examples of data viz use provided and much less apparent than it was in the short survey responses. The underlying *efficiency* argument can be detected in comments such as the data viz allowing clients to “clearly extrapolate the information that they wanted” [FF-10], but *efficiency* is not heavily emphasized in the specific examples of data viz use.

U-FE Framework. As expected, discussion of data viz use consistent with the *U-FE Framework* is *very apparent* when analyzing the specific examples of data viz *use*. Twenty-six (89.66%) of the use cases reference this model. Although conceptualizations of *use* consistent with the *U-FE framework* were not apparent in the shorter survey responses, once engaged in discussion during the interviews, the evaluator emphasis on utility was apparent in their specific examples of data viz *use*. Interestingly, specific terms such as *instrumental*, *conceptual*, or *symbolic use* were never used, but their descriptions of *use* often suggested these categories. For example, a clinic adopting the three recommendations indicated as easiest to adopt by the visual [AA-10], using the visuals to advocate for more funding [TT-8], government officials presenting data viz in reports [GG-10], or making revisions to the agency's website [YY-5] could be examples of *instrumental use*; a client indicating "we learned from this" [KK-10] or "remembering key takeaways" [MM-7] could be examples of *conceptual use*; and the numerous examples of using the colors and style of data viz to support client brand could be examples of *symbolic use*.

Explain or Explore Model. The *explain* ←→ *explore model* is *apparent* when analyzing the specific use cases, but as with the *efficiency rationale*, it was less obvious here than it was in the short survey responses. Cases frequently reference the audience receiving the visual, but then suggest uses beyond simply *explaining*. Fifteen (51.72%) of the specific examples were coded to the *explain* ←→ *explore* model with emphasis on the *explain* end of the continuum. Only one of the examples, a dashboard project where the client continued building and maintaining the dashboard after its creation, explicitly discussed a client using data viz to *explore* data stating, "now they're doing all these really cool analyses in Tableau" [WW-6].

Data to Insight Model. The *data* → *insight model* was only *slightly apparent* in the specific examples of use with only 12 (41.38%) cases referencing it. Examples of use such as "it went to the board and they looked at it and were like, 'Wow'" [BB-15]; clients engaging in "informed decision-making" using the data viz [GG-10]; clients "showing understanding" [EE-4]; and clients engaging in sense-making and asking more questions [LL-4] could all be examples of *use* consistent with the *data* → *insight model* even though the model itself was never mentioned.

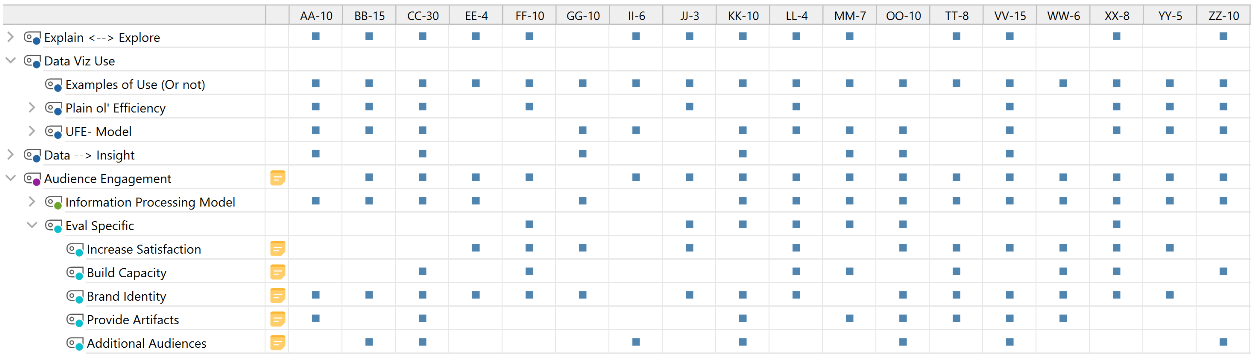
Audience Engagement. The second study question regarding whether there are *other reasons for using data visualization suggested by this sample of practicing evaluators* again suggested the importance of audience apparent in the survey responses. Whereas, in the short survey responses, a simple emphasis on audience was apparent, within the specific examples of data viz *use*, a theme of *audience engagement* emerged. Data viz was conceptualized not just as a way to *explain* data to an audience, but as a way to *engage* that audience by attracting their attention [BB-15; WW-6], getting them excited and holding their attention [EE-4], increasing satisfaction with evaluation [FF-10; LL-4], and providing data viz artifacts that give the audience the ability to repurpose data viz to engage additional audiences [TT-; ZZ-10]. These conceptualizations of *audience engagement* emphasized data viz *use* beyond simply *explaining* data to an audience.

General Conceptualizations of Data Viz Use

In addition to an inspection of the specific examples of data viz use provided, each interview was re-read in its entirety and analyzed for conceptualizations of data viz *use* whether related to a specific example or not. The *a priori* models again served as sensitizing concepts. In this analysis, evaluator emphasis on *utilization* and *audience engagement* emerged even more strongly than within the specific examples of data viz use. All models could be discerned in at least 50% of the interviews, see Figure 10, so the definition of *apparent* relies heavily on my perception of the effort required to extract the model from the data and the extent to which interviewees are describing the model (or components of the model) in alignment with the terminology as it is used in literature. See methods section, Table 1, and Table 3 for more details.

Figure 10

Summary Grid Demonstrating That All Models Can be Discerned in a Majority of Interviews



Note. This screenshot is taken from MaxQDA showing that the models could be discerned but does not note the relative ease with which each could be identified.

Increased Efficiency. The rationale that data viz *increases efficiency* over text-based communication is *very apparent* in interview responses when speaking about the general value of data viz. The specific word *efficiency* is used in most of the interviews with an emphasis on effective communication. *Efficiency* is also discussed using related terms such as, “concise and clear” [BB-15], “better digest information” [CC-30], “simple, clear” [FF-10], “easily and clearly” [JJ-3], “easier to see” [XX-8], and “easy for folks to understand” [YY-5].

U-FE Framework. The *U-FE framework* is *very apparent* when analyzing interviews in their entirety. While the *U-FE framework* is still not mentioned by name, the evaluators clearly demonstrated a *utilization* focus consistent with the *U-FE framework* during the interviews. Several evaluators even indicated that they defined the example work their *best work* specifically because it was *used* with comments such as, “I chose this one because the community really liked it, and it was useful” [BB-15], “I picked these ones because they have been used to make decisions” [KK-10], and “I picked these in part because they've been widely used … I think about how I was defining best in my mind, that was part of it. I think there are more elegant dashboards I've designed, but that went much less used” [XX-8].

Relatedly, the *U-FE framework* conceptualization of *use* was emphasized as being more important than the artistic value of the data viz in many of the interviews with comments such as, "I can make really fancy [data viz] but my definition of best … is not necessarily the glitziest or the most professional looking viz. It's about viz that makes people make different decisions" [KK-10]. Different emphases on the role of art in data viz was noted between evaluators and non-evaluators by Douville and colleagues (in press), with non-evaluators more likely to emphasize the artistic role of data viz. Consistent with that finding, in this analysis, evaluators clearly leaned toward function over form with statements such as, "if the most effective form is going to be a bar chart, then that's what I will strive to use. I want something that is going to be easy for folks to understand so that they will be better informed to make decisions" [YY-5].

An emphasis on *instrumental use* is apparent in many interviews, particularly focused on the importance of using data viz for programmatic decision-making [GG-10; KK-10] and funding decisions [II-6; VV-15]. Many interviewees emphasized the value of data viz in increasing understanding and making people think differently about data. These could be considered *conceptual use* as future decisions could be impacted by that understanding. Many evaluators spoke of using data viz to support client branding decisions and because clients expect to see data viz. These could be considered examples of *symbolic use* and are also very consistent with the audience focus emphasized by evaluators.

Explain or Explore Model. The *explain* ←→ *explore model* is also *very apparent* in the interviews. As noted previously, using data viz to *explain* to an audience was the most common *use* described in the surveys and is also present within most interviews particularly in the context of using data viz to report findings to an audience. However, the overlapping nature of these conceptual models was also strongly apparent here. Many interviewees appeared to use the words *explain* and *storytelling* synonymously or used both terms in their descriptions. And both of these conceptualizations were clearly centered around doing so for an audience.

Whereas most discussion along the *explain* ←→ *explore model* emphasized the *explain* end of the continuum, analyzing the interview responses outside of the specific examples of *use* yielded examples of evaluators and their audiences using data viz to *explore* data. Much of the *explore* discussion was focused on using data viz to see or "reveal" [MM-7] patterns in data, particularly in participation data [KK-10; LL-4], or to better understand group comparisons not apparent in average calculations [MM-7]. Most of the *explore* comments referred to the evaluator, not the client, using the data viz to *explore* the data to find answers to questions with comments such as, "if you don't know what your answer is, the next step is exploring the data" [II-6], or "sometimes, the takeaway is not clear until you put it into a certain format" [MM-7], but one evaluator described creating and providing data viz to the client, then the client *explored* the data viz and determined "the key finding" [BB-15]. Several evaluators discussed using data viz to determine what the questions should be "because we don't even know the questions we're asking at this point" [KK-10].

Without naming the *explain* ←→ *explore model* directly, several evaluators spoke to the cyclical nature of creating and refining a data viz within an *explain* ←→ *explore* context whereby *exploring* the data viz creates new questions that then need to be visualized and *explained* to others. One evaluator provided a very nuanced example in which the evaluator creates data viz to describe complex qualitative project requirements and workflows for government evaluation projects because "I can't tell you how many government proposals are written out kind of poorly in the ways of how the different elements of the proposal interact with each other" [AA-10]. The evaluator went on to describe the cyclical process of creating the data viz including deeply exploring qualitative aspects of the proposal, drafting and redrafting the viz, trying to understand requirements and connections using the viz, and then using the viz to *explain* this understanding to others: "And that graphic usually ends up front and center at the beginning of every proposal" [AA-10].

Data to Insight Model. The *data* → *insight model* is denoted as only *slightly apparent* in the interview data. Although the model was never referenced by name, components of the model can be discerned in nearly every interview. Only three of the interviewees used the word "insight" related to data and only one of those in a manner that indisputably represents the model, saying that they use data viz to "help the end users get oriented and more comfortable working with more sophisticated data visualizations that ultimately probably enable them to gain insights that they wouldn't otherwise gain from a bar chart" [YY-5]. However, nearly every interview discussed using data viz to increase "understanding". Examples of data viz *use* such as clients engaging in *"*informed decision-making" [GG-10] or "enhanced action" [MM-7] using the data viz, or clients engaging in sense-making and asking more questions [LL-4] could all be examples of *use* consistent with the *data* → *insight model*. Arguably, these examples can be applied to other models presented here, as well, which is consistent with the overlapping nature of these models. There is also support for the *data* → *insight model* within comments such as "it went to the board and they looked at it and were like, 'Wow'" [BB-15] which could be associated with the epiphany or "aha moment" referenced in the model.

Audience Engagement. The second study question regarding whether there are *other reasons for using data visualization suggested by this sample of practicing evaluators* found additional support and detail for a model of *audience engagement* when the interviews were re-analyzed in their entirety. The four areas of *audience engagement* identified in the specific examples analysis were also apparent in the interviews: attracting audience attention, getting them excited and holding their attention, increasing satisfaction with evaluation, and providing data viz artifacts that give the audience the ability to repurpose data viz to engage additional audiences. These subthemes of *audience engagement* are identified as (1), (2), (5), (8) and (9) in Figure 11. Four additional subthemes around (3) *encouraging interaction*, (4) *aiding memory or learning*, (6) *building capacity*, and (7) *using client brand identity* became more apparent through the analysis of full interviews, as will be discussed below.

Figure 11

Areas of Audience Engagement Emphasized by Practicing Evaluators

Subthemes (1) and (2) in Figure 11 are relatively straight-forward and rely heavily on the basic tenants of human biology and data viz to (1) *attract attention* and then (2) *hold attention* for deeper engagement. The goal of initially attracting attention, or “eye-catching” [EE-4], was described by one interviewee as being done through using, “fundamental design principles, the basic pre-attentive attributes, and how you leverage and use those effectively,” pointing out that “all those tiny, small choices actually make a big difference” [XX-8]. After the attention is (1) grabbed using the data viz, interviewees discussed the importance of (2) holding that attention and keeping the audience engaged beyond the initial glance. Terms such as focus, attention, and thinking were used in this context with an emphasis on keeping “people's eyes on the screen and to really get at sort of these more difficult concepts” [VV-15]. Although the overarching relationship appears directional from (1) to (2), there were hints in the interviews that (1) *attracting attention* is not a single event such as having data viz on the cover of the report to get people to open a report, rather that it must happen repeatedly to keep people engaged. This supports using data viz throughout a report as is summarized well by one interviewee:

> Every now and then, you'll notice a big number standing out, that's part of data viz, there are some icons, and in the hopes that this report would be a lot more compelling and readable than if you were just looking at black and white Word document pages. So, hopefully this provides something that is inviting and people will want to read. And then of course, in evaluation, we want our evaluation data to be used. [OO-10]

Interviewees described various benefits of holding that attention, which are categorized here as subtheme (3) *encouraging interaction or connection*, and subtheme (4) *aiding memory or learning*. The relationship between these two subthemes is not completely clear from the discussions. These two subthemes may overlap (2) *holding attention*, immediately follow (2) *holding attention*, or may create a feedback loop as continuing (3) *interaction or connection* makes (2) *holding attention* possible longer and allows the processes that (4) *aid memory or learning*.

Subtheme (3) could describe using data viz to *encourage interaction* with the evaluator [ZZ-10] such as asking more questions or encouraging interaction with other participants such as engaging in discussions. One participant described this as, “moving the conversation forward” [BB-15]. The perceived value of data viz in (1) *attracting attention* and then (2) *holding attention* to (3) *encourage interaction or connection* in this context, is summarized as:

> I wanna be able to sit in the meeting and throw this up on the screen and put it in their hand and get them to focus, 'cause it's hard to get people to read things. Hence the pretty pictures. And then [I can say], ‘You want me to go through the different pieces of it?’ [ZZ-10]

Another way that (3) *interaction or connection* was discussed was in terms of an internal process within the audience in which they connected with the data in a meaningful way such as taking “ownership” of the data [YY-5], or changing goals based on their new understanding: “[They wanted to] learn more and do better and in fact, I got emails back like, ‘We're setting new goals, this isn’t what we want’” [KK-10].

Subtheme (4) *aiding memory or learning* was most frequently discussed in terms of learning with emphasis on using data viz as another tool to teach people with different learning styles [AA-10] or as part of a learning process to help clients organize their thinking just as the visual organizes the content [BB-15]. Only two participants made specific mention of memory in “making a difference on people remembering key takeaways” [MM-7] and that “the human brain can only latch on to or remember like a couple of things from a presentation” [VV-15], but both terms are named in this model because memory and learning are so closely linked in related models.

Subthemes 5, 6, 7, 8, & 9 are heavily inter-related, but the perceived relationships to each other and the first four subthemes cannot be established in this research phase. The descriptions that follow are based on the secondary analysis of existing data method used in this research phase. It is anticipated that a more complete model of *audience engagement* will be described and established in a subsequent research phase.

Subtheme (5) concerns the emphasis nearly all interviews placed on the importance of *client satisfaction* with data viz. A wide variety of assumptions regarding the role of data viz in increasing satisfaction were presented. Sometimes, the description could almost overlap with (1) *attracting attention* such as using, "[data viz] as a tool to gather [client] enthusiasm or their engagement early on in the process" [FF-10] or overlap with (3) *interaction and connection* in getting the audience to connect with the data such as to, "get buy-in from people who are kind of dismissive of evaluation and skeptical of evaluation, and they just are indifferent to what you do" [FF-10]. There were also suggestions that data viz was associated with evaluator credibility and expertise: "I wanted it to be slick. And I wanted them to know that I know what I'm doing, and I am an expert in my field" [GG-10]. As well as suggestions that using data viz poorly can harm *use* and credibility: "They have way more impact when it's not just tons and tons of colors all over the place. People don't take it as seriously. But then when you can use it with intention, I think it can do a lot for you" [WW-6].

Perhaps unsurprisingly in interviews with experienced evaluators, discussion of working to (6) *build client evaluation capacity*, including the ability to understand and create data viz, was a frequent topic in the interviews. Many interviewees discussed the data viz they produced as "not just a product, but also a capacity development opportunity" [XX-8] with one stating, "most of my work is capacity building work, and everything that I create for them, I want them to be able to replicate and understand" [TT-8]. Another specifically indicated, "it's like data viz capacity building" [ZZ-10].

The subtheme (7) *branding* is significant in that every interviewee discussed their experiences with and rationales for using the client's logos, colors, and style as they created data viz and report layouts. Only one participant, working at an international evaluation consulting firm, used their own firm's branding exclusively in data viz design. All others chose to use either colors and images significant to the stakeholders or the specifically documented client's style guide, with one participant even calling a particular data viz experience "very much a branding exercise" [LL-4]. Evaluators used the client's style guide either because they were required to, which was often the case, or because they felt using the client's colors and style would support other aspects of the *audience engagement* model. For example, using data viz to (3) *connect* the audience to the data through the data viz was described as:

> We always use their color scheme. …We figure if we use their colors, then when they see it, they'll identify with it more. And they'll think like right, this is something internal, this is something from us, this isn't just something from some random group trying to tell me what's going on. So, helping them identify with it by using their own colors. [EE-4]

In fact, the evaluators interviewed prioritized using the client's branding even when doing so led to trade-offs in design appearance or accessibility. For instance, they often discussed not preferring the client colors used because they were not "aesthetically pleasing" [TT-8], or had meanings associated with the colors that were not appropriate for the intent of the viz such as, "blue doesn't necessarily feel like an alerting color" [XX-8], or with comments such as, "The choice of colors is matching the client's logos. They wouldn't be my first choice of colors... I kind of think that beige is a bit hard to read in that" [GG-10]. Concerningly, many evaluators discussed having to trade-off making the visuals color-blind friendly and accessible when using client colors. Some described slightly altering client colors to meet accessibility requirements but not being fully satisfied with the results.

Subtheme (8) *providing artifacts*, refers to evaluators creating data viz materials such as report graphics, presentation slides, and hand-outs that could be used independently from the evaluation process and report. The word *artifact* was not used by any of the interviewees to describe these stand-alone data viz products and there was no sense from the interviews that the participants view this as a separate activity. This subtheme is strongly present in the interviews and thoroughly integrated into the other subthemes, but its relationship(s) to the other subthemes was not clear from this interview data. What was clear is that the creation of these data viz artifacts is heavily entwined with subtheme (7) *branding* in that evaluators often discussed the importance of providing products that were consistent with the client's brand identity. Further, subtheme (8) *providing artifacts* was most closely linked not to artifacts that the client would use for their own purposes, but to creating materials that clients could then provide to (9) *additional audiences*.

The emphasis on subtheme (9) *additional audiences*, was very apparent throughout the interviews, from a general acceptance that – once created and provided to a client – the evaluator would have little control over how a data viz was used, to some evaluators only producing data viz for distal use: "I very rarely make visualizations that I do anything with. I create them for people to share with their stakeholders" [KK-10]. In many of these discussions, individual data visualizations were discussed as if they were individual artifacts or vestiges – although those words were never used – that could exist separately from the context in which they were originally used.

One frequently mentioned concern with this loss of context is that the evaluator cannot always know the potential distal audience for the data viz. They discussed the many design challenges associated with not knowing the eventual medium (e.g., paper hand out, slide presentation, website, etc.), nor the distal audience (e.g., funders, community members, government regulators, etc.) of the data viz, as discussed by Douville and colleagues (in press). Although frustration was sometimes discernable, the majority of participants appeared to consider secondary use of data viz for distal audiences simply a component of being responsive to client needs. Some presented this as a positive factor in helping their clients to, “get some local press” [VV-15] or use the data viz as “marketing materials in addition to just the standalone report” [WW-6].

Many of the evaluators interviewed elaborated on the extents to which they go to try to understand how the client will use the data viz with secondary audiences: “I wanted to do something 8 1/2 by 11 … we would wanna print this out and literally hand it to some of the trainers and other stakeholders, because that's just what the audience tends to like” [ZZ-10]. Others emphasized keeping the data viz appropriate for multiple audiences: “Next time, we're not writing a report, we're doing a power point presentation that they can share whatever slides are appropriate for whatever audience” [TT-8].

One person clearly connected (6) *capacity building*, (7) *branding*, and (9) *additional audiences* to each other by emphasizing teaching clients to use branding:

> Especially when we're talking about the mission of some of these smaller, community-based organizations, who really, marketing's the furthest thing from their mind, because they're on the frontline. This is a small way to even get them thinking about how important consistency is. [WW-6]

Phase I: Discussion

The purpose of this Phase I study was to determine whether there is support for the *efficiency* rationale of data viz and/or any of the three models derived from the literature (*U-FE framework*, *explain* ←→ *explore*, and *data* → *insight*) in a sample of practicing evaluators. If support could not be identified, there would be no reason to present that model in Phase III. This analysis was also seeking any conceptualizations of data viz *use* not fully explained within the three models; a model emphasizing data viz *use* for *audience engagement* was uncovered. All models were present to various extents in each method of analysis, but there were interesting patterns in the extent to which they were present in each, see Figure 12.

Figure 12

Appearance of Each Model by Method of Data Analysis

Question	Short Survey Response	Interviews: Specific Examples	Interviews: General Discussion
1(a): increased *efficiency*	apparent	slightly apparent	very apparent
1(b): *U-FE framework*	slightly apparent	very apparent	very apparent
1(c): *explain* ←→ *explore*	very apparent	apparent	very apparent
1(d): *data* → *insight*	apparent	slightly apparent	slightly apparent
2: other reasons for data viz *(audience engagement)*	apparent	very apparent	very apparent

Note. Density of colored bar indicates how apparent each model was by method of analysis; categories are *very apparent*, *apparent*, and *slightly apparent*.

Efficiency Rationale is Generally Accepted

It is *apparent* that evaluators are aware of the *efficiency* rationale of data viz as they referenced it frequently in their short survey responses and within the general discussion. The strong emphasis on this reason for using data viz within the short survey responses demonstrates that, when given only a brief opportunity to provide a reason for using data viz, this rationale was quickly evoked. And, when engaged in general discussions of evaluation *use*, the *efficiency* rationale is frequently mentioned. However, the noticeably decreased emphasis on this rationale within the specific examples – when conceptualization of *use* within an evaluation context is being specifically addressed – suggests that evaluators were considering reasons other than *efficiency*. This suggests that *increased efficiency* is a dominant and accepted underpinning rationale for data viz *use* but is not conceptualized as the ultimate reason for using data viz in evaluation.

The U-FE Model is Extended to Data Viz

The *U-FE* model is only *slightly apparent* in the responses to the short survey question but is *very apparent* within the specific examples and general discussion. This stark difference is likely because the survey question asks about data viz but does not mention evaluation nor ask participants to respond from an evaluation standpoint. This framing elicited the fundamental “to increase *efficiency*” value proposition of data viz. The *U-FE* model was not specifically described to the interviewees and they were not asked to consider their data viz *use* within that model, explaining the very infrequent direct reference to the model itself. Nevertheless, when discussing evaluation and when talking specifically about evaluation *use*, a traditionally accepted evaluation-based conceptualization of *U-FE* was very apparently extended to data viz.

When asked to provide examples of *use*, participants tended toward trying to provide examples of *instrumental use*, as is consistent with the historical emphasis on *instrumental use* over other types of *use* in evaluation. Some interviewees appeared almost apologetic that they couldn’t provide *instrumental use* examples with statements like, “In terms of how they influence decisions, I don't know. But I think in terms of dissemination, or use, or visibility, yeah, definitely examples of that. Again, does that translate into action? That's I think sort of a stickier question” [II-6] or “I would love to have a story where someone comes back and said, ‘You know, that pie graph really made us think about this aspect of the program and make changes.’ But, you don't often get that” [OO-10]. One participant gave a very clear example of data viz use in which the client made program decisions because the data viz "showed them what were the easiest things to do" but then still went on to lament not having grander examples of *use* by saying, "I can't think of anything off the top of my head where major policy decisions have been made off of a visualization I did" [AA-10].

The Explain ←→ Explore Model is Apparent

The *explain* ←→ *explore* model which – as demonstrated in the literature review – is common within evaluation, follows a very similar pattern to the *increased efficiency* rationale. The *explain* ←→ *explore* model is clearly referenced in the short survey and within the general discussions, but is less emphasized within the specific data viz examples. Perhaps, as with the *efficiency* rationale, this widely accepted and discussed rationale is simply accepted as an underlying or obvious rationale for data viz *use*.

Evaluators using data viz for their own purposes (e.g., to *explore* the data) was not mentioned in any survey responses and was not apparent in the specific examples of data viz use. However, this particular aspect of the *explain* ←→ *explore* model could be discerned in the general analysis of the interviews, even though a specific question about their own use of data viz was not asked. This suggests that the evaluators do use data viz for their own *explore* purpose, but do not formally name it "use." This could simply be because data viz literature and training materials within evaluation emphasize the *explain* aspects of data viz for an audience, so the evaluators had not considered their own use of data viz as data viz. This would be consistent with the heavy audience focus they demonstrated throughout the interviews; perhaps they conceptualize use from the audience point of view, not their own. Not emphasizing these internal types of use would also be consistent with their *U-FE* preference toward providing clear-cut examples of *instrumental use* over more abstract, less defensible, examples of *use*.

Not framing *explore use* as a type of *use* could be a result of simply not having considered it before, or it could be that they actually do not conceptualize their own *use* as *use* due to their heavy audience focus. Limiting conceptualizations based on naming conventions are found in other examples of activities that could be named data viz, but are not. For example, no interviewees described logic models as data viz even though they could easily be considered a form of data viz. This also introduces the possibility that evaluators are using data viz to *explore* data more often than is apparent because they are simply not identifying it as such.

The Data → Insight Model Can Be Discerned, With Effort

The *data* → *insight* model was never mentioned in the data set. This was an expected finding as the literature review demonstrated that this is not a popular conceptualization of data viz in evaluation. However, the model is still marked as *apparent* or *slightly apparent* in each of the analyses because important components of that model were very apparent in the data. Many of the interviews described components such as using data viz for decision-making, storytelling, information, supporting knowledge, and often discussed the "wow" factor of data viz that is associated with insight. The alignment with key components of the *data* → *insight* model, without ever naming the model, suggests the importance of presenting this model to evaluators to see if they find it a useful conceptualization once it is named and explained to them.

An Audience Engagement Model is Very Apparent

While *efficiency* and all three models can be discerned to varying extents within the data, the most apparent emphasis across all data and analysis methods is the *audience engagement* model. This model was not uncovered in the literature review and has not been emphasized as a data viz model in evaluation.

Evaluators were very focused on the audience(s) for their data viz. This emphasis is consistent with their translational role in communicating knowledge to stakeholders (Valéry, 2007) and with the reality that clients must be satisfied with the product and service to secure future contracts. This emphasis on audience should be expected given the very nature of evaluation as an applied field. Unlike traditional research, evaluation is conducted for a purpose and in response to an external need (Alkin & King, 2016).

Arguably, subthemes (1) *attract attention*, (2) *hold attention*, (3) *interact (connect)*, and (4) *remember (learn)* describe benefits of data viz that are applicable to most fields. They may simply be an audience centric way of describing the underlying cognitive processes expected to occur when using data viz. Indeed, if we consider (3) *interaction* as describing the working memory process and (4) *learning* as long-term memory, then subthemes (1) through (4) are reminiscent of the information processing model of memory as described by Huang et al. (2009, p. 3) in Figure 1 and supports cognitive load theory (Sweller et al., 2019). As the *efficiency* rationale is closely tied to cognitive load theory, subthemes (1) through (4) may simply be an audience centric conceptualization of the *efficiency* rationale.

Framing (3) *interact (connect)* as an internal process has strong overlap with the decision-making process within the *data* → *insight* model and with *instrumental use* within the *U-FE* model. The emphasis on the importance of (4) *learning* is not unique to this model. Learning can be considered a component of *conceptual use* in the *U-FE framework* and within the *knowledge* category of the *data* → *insight* model. The role of data viz in assisting learning and memory is consistent with the brain chemistry changes associated with the *data* → *insight model* as well as research suggesting that embellished data viz may be more memorable (Batch, 2018) or persuasive (Rykaczewska, 2021).

What is new in this analysis is not the categories nor explanations of how the data viz is expected to benefit communication; the understanding gleaned from this analysis is that these evaluators tied their conceptualization of data viz use with audience. Even the subthemes that could apply to any field were described from an audience centric point of view.

Subthemes (5) *client satisfaction*, (6) *build capacity*, and (9) *additional audiences*, along with the related subthemes of (7) *branding* and (8) *providing artifacts* are the most evaluation specific aspects of this emerging *audience engagement* model. The emphasis on (5) *client satisfaction* is self-explanatory, but the heavy emphasis on (9) *additional audiences* can be tied to evaluator desire to extend the usefulness and impact of the evaluation generated knowledge to additional audiences and also to a strong desire to support evaluation capacity building. Many evaluators emphasized wanting to give their clients the interest and the tools to do data viz work themselves. They also expressed a desire to give their initial audience the ability to repurpose their visuals and extend them to distal audiences. Related to this capacity building focus was an explicit desire to produce data viz as artifacts that extend their usefulness beyond the initial dissemination and can be used to engage distal audiences.

Phase I: Limitations and Considerations

The main limitations of Phase I are related to the secondary analysis of existing data method used, which created a mismatch in terminology around *use* and did not allow for a thorough description of the emerging *audience engagement* model. Another method used in this study, analyzing data based on specific examples of use, was of limited value, but has an unanticipated benefit as noted below.

Terminology Around Use

While it is convenient to employ a secondary analysis of existing data for this study, relying on data not collected to specifically answer the current research questions presents a slight mismatch in terminology around *use*. Regarding the survey data, *use* was framed as *purpose*, which may have encouraged against a utilization focused definition of *use* in the responses. Indeed, the lack of *U-FE* specific definitions was noted in the findings. While potentially biasing participants away from a *U-FE* specific definition, the imprecise terminology of *use* and *purpose* is not inconsistent with the present study's open-ended definition of *use*. Some arbitrariness on the definition of *use* going into this study should – hopefully – allow richer understanding of participant conceptualizations of data viz *use* to emerge from within the evaluation context.

Relatedly, the original data collection interviews often switched from a standardized open-ended interview format to "responsive interviewing" wherein follow-up questions and probes were altered based on the flow and content of the interview (Patton, 2015). As noted in Appendix D, the researchers did not specifically ask the *use* question of three participants. While encouraging deep and interactive interviews was a goal of the original research team, the resulting variability in response sets is a limitation for this secondary analysis. Related to this inconsistent presentation of the interview questions, another potential limitation is that the interview tool included a biased definition of *use* by parenthetically describing *use* as "*(brought up, mentioned)*" rather than providing participants with either an open interpretation of *use* or a *U-FE* definition of data viz *use* appropriate to evaluation. A review showed that half (*n* = 9) of the interviews [BB-15; CC30; FF-10; JJ-3; KK-10; LL-4; MM-7; WW-6; YY-5] specifically included this parenthetical description.

Secondary Analysis of Existing Data Method was Insufficient to Fully Describe the Model

This study relies on secondary analysis of an existing data set of survey responses and interviews. This rich data set of eighteen surveys and interviews was more than sufficient to address the research questions as far as confirming the presence of the models. It even supported a greater understanding of the prevalence of each model within the sample. Importantly, this secondary analysis of existing data method was useful in clearly supporting that an additional model of *audience engagement* exists within this sample of practicing evaluators.

While this data set was not as ideal as would be expected from a data set designed and collected to answer these specific questions, it was more than sufficient for an introductory study to determine the existence of the specified models and potentially uncover other conceptualizations of data viz *use*. I am indebted to the participants for their time and to the original research team for their efforts in collecting, organizing, and preserving this data.

However, the inability to ask follow-up questions during the interview process to better understand the relationships between the subthemes of the emerging *audience engagement* model, prevented a thorough description of the model. While the goal of this study was not to create a grounded theory of data viz *use* in evaluation, the emerging *audience engagement* model is relevant enough within evaluation to warrant further analysis and description. As will be discussed in the conclusion, Phase II will be added to this research study to further describe this model.

Benefits and Drawbacks of Identifying Specific Examples of Use

One method selected for this study, identifying and then analyzing specific examples of data viz *use* from the interviews, was implemented to maintain consistency with the original study which focused heavily on the context of data viz design in evaluation. That approach had limited usefulness for this study as it did not contribute meaningfully to understanding the conceptual models. It did serve to highlight the *U-FE* focus of the evaluators and make that model more apparent, but this small contribution is outweighed by the effort of engaging a third analysis approach for each model. The findings were well supported by the other two methods, such that the specific examples analysis could have been excluded from this dissemination. However, there was one unanticipated finding that provides some small contribution to understanding the prevalence of examples of data viz *use*.

This unanticipated finding was the relatively large number of examples of data viz *use* provided. It has been long lamented within evaluation that there is a dearth of examples of evaluation *use*, but this sample of 18 evaluators provided 29 examples of data viz use. This represents more specific examples of data viz *use* than I was expecting to identify, especially given that use was not the primary focus of the 2021 interviews. Other researchers may argue that some of the examples provided do not constitute *use* in a purist *U-FE* conceptualization of *use*. It is true that this analysis was based on an intentionally wide interpretation of *use* to better understand participant conceptualizations. As some examples may not stand up to scrutiny, all are presented in Appendix D so that others can contribute their own perspective on which constitute *use* in the context of data viz.

Phase I: Conclusion

Phase I explored evaluator conceptualizations of data viz use from interviews with eighteen evaluators and determined that the *efficiency* rationale of data viz as well as the three models derived from the literature (*U-FE framework*, *explain* ←→ *explore*, and *data* → *insight*) are present in this sample. A fourth conceptual model emphasizing using data viz for *audience engagement* was uncovered and partially described. Further research (Phase II) is needed to fully describe this emerging model of *audience engagement*. Further research (Phase III) is also needed to better understand whether and how evaluators find each of these conceptualizations of data viz use beneficial and useful to practice.

Efficiency Rationale

It is apparent that the *efficiency* rationale of data viz use is a dominant conceptualization and an accepted underpinning rationale for data viz *use* in evaluation. However, this analysis suggests that evaluators also conceptualize other reasons – beyond efficiency – for using data viz in evaluation. Better understanding these other conceptualizations and their relationship with the *efficiency* rationale will require further research.

U-FE Model

Evaluators in this sample appeared to extend the *U-FE* framework of evaluation use to data viz, without directly referencing the model itself. As the original research protocol did not specifically address the *U-FE* model, further research will be needed to understand whether evaluators consider the *U-FE* model helpful in conceptualizing data viz *use*.

Explain ←→ Explore Model

The *explain ←→ explore* model is common and accepted within this sample and, like the *increased efficiency* rationale, appears to be simply accepted as an underlying or obvious rationale for data viz *use*. Further research will be needed to tease out whether this is a stand-alone and useful conceptualization within evaluation or if it is merely a component of the other conceptualizations.

An additional consideration around conceptualizations of data viz *use* emerged through the analysis of the *explain ←→ explore* model that bears further study. Evaluators did give examples of *using* data viz for their own purposes, but they very rarely described their own *explore use* of data viz as *use*. As clarifying questions were not asked, it is unknown whether they simply have not had prior occasion to conceptualize their own *explore use* as a form of data viz *use* or if they specifically do not consider it *use*. There could be tremendous benefit to expanding the conceptualization of data viz *use* in evaluation to include the evaluator's *explore use* because most of the underlying tenants of data viz that benefit *efficiency* gains can be harnessed for internal *explore use* easily. This same argument holds for data viz of qualitative data.

Data → Insight Model

The interview data appears to suggest that evaluators conceptualize data viz consistent with key components of the *data* → *insight* model, without ever naming the model. However, there are alternative explanations for my identification of components of the *data* → *insight* model within the interviews. Perhaps all the models have so much overlap that nearly any statement could be applied to any model. Or, perhaps the *data* → *insight* model resonates with me and, due to my bias, I have inappropriately applied this model to the data. As there is no suggestion in the literature nor through this secondary analysis of existing data that this model is recognized nor useful within evaluation, further research will be needed to ascertain whether evaluators find this framework – to tie together concepts such as knowledge building and storytelling that they already consider in their work – appropriate and helpful.

Audience Engagement Model

An evaluation specific data viz *use* model considering *audience engagement* was clearly present within the data, although it is not yet fully described. There is tremendous overlap between the three models presented and the emerging *audience engagement* model, as well as clear overlap between the *audience engagement* model and the information processing model contained within cognitive load theory. As the model is similar enough to existing models to make rational sense, but also specifically evaluation focused enough to potentially benefit evaluation, there is justification to conduct further work describing this model and then present it to evaluators to understand whether they find this conceptualization helpful and useful.

Future Direction

The eventual goal of this project (Phase III) is to present the *efficiency* rationale of data viz, the three models derived from the literature (*U-FE framework*, *explain* ←→ *explore*, and *data* → *insight*), and the newly identified model of *audience engagement* to a sample of professional program evaluators to determine whether and how each of these conceptual models are useful in evaluation. While the *efficiency* rationale and the three models derived from the literature are sufficiently described for this purpose, the *audience engagement* model emerged from a secondary analysis of existing data and is not described well-enough to present to other evaluators.

Consistent with an exploratory study, it was unknown at the beginning of Phase I that a conceptual model of *audience engagement* would emerge from the secondary analysis of existing data. The *directed content analysis* (Hsieh & Shannon, 2005) approach employed in Phase I was a sufficient method to identify the existence of the conceptual model of *audience engagement*, and even sufficient to identify the subthemes noted; but it was insufficient to describe the concepts well enough to explain the connections and relationships between the subthemes.

Therefore, Phase II will use follow-up interviews to better understand the perceived theoretical relationships between the subthemes within the emerging *audience engagement* model.

Phase II

Phase I relied on a secondary analysis of existing data to determine whether the underlying *efficiency* argument for data viz *use* as well as the three proposed models of data viz *use* were present within a sample of practicing evaluators who are experts in data viz. Findings suggest that the *efficiency* rationale and all three models (*U-FE Framework, explore* ←→ *explain*, and *data* → *insight*) were present to varying extents, see Phase I and Table 3 and Figure 12 for details. The *efficiency* rationale and these models will be articulated into explainer videos during this phase.

A second Phase I question asked: *Are other reasons for using data visualization suggested by a sample of practicing evaluators?* All three routes of analysis used in Phase I suggested that an evaluation specific data viz *use* model considering *audience engagement*, with related subthemes, is present in the data. Although this model emerged in Phase I, see Figure 11 in Phase I, it could not be fully described because the data set and the analytical method used were insufficient. As the model is similar enough to existing models to make rational sense, but also specifically evaluation focused enough to potentially benefit evaluation, there is justification to conduct further work describing this model and then present it to evaluators to understand whether they find this conceptualization helpful and useful.

The discrepancy between the apparently conflicting statements made so far, that data saturation was both achieved and not achieved in Phase I, can be explained by the various ways in which the term "data saturation" is used. The term *saturation* originated in grounded theory as "data saturation or thematic saturation and refers to the point in data collection when no additional issues are identified, data begin to repeat, and further data collection becomes redundant" (Kerr et al., 2010). The emphasis was on "theoretical saturation" to mean that the theme was thoroughly explained and was not concerned with sample size. As the term *saturation* was applied to other qualitative research approaches, it has been construed to have a variety of meanings and has lost clarity around determining when it has been achieved (Kerr et al., 2010). Saturation became more closely associated with justifying that an appropriate sample size has been achieved and less concerned with the original definition concerning theoretical saturation (Hennink & Kaiser, 2022).

In Phase I, the often-used definition of *saturation* as "data saturation," "code saturation," or "thematic saturation" (Hennink et al., 2017) was met through analysis of the interviews in that no new concepts, issues, or themes were emerging as the last of the interviews were analyzed. All of the subthemes were present in multiple interviews. What was missing was the more difficult to obtain definition of saturation needed to thoroughly describe a theory: "the theory is dense and logical and there are no gaps in the explanation" (Corbin & Strauss, 2015, p. 140). This is what Hennink and colleagues (2017) differentiate as "meaning saturation." While the concepts were consistent, there was not sufficient development of the properties and dimensions of those concepts (particularly *branding* and *satisfaction*) to explain how the interviewees conceptualized the connection and relationship between those concepts and the model as a whole. In short, the subthemes were obvious, but I could not describe how the evaluators interviewed feel the subthemes contribute to *audience engagement*.

Phase II was added to this study to allow sufficient description of the emerging *audience engagement* model to present it to practicing evaluators in Phase III. This phase was not intended to be a stand-alone study; it was added only to describe the conceptual model well enough to articulate it and present it for review in Phase III. Once this *audience engagement* model was sufficiently described for this purpose, it will be articulated into a short explainer video to be used alongside the other models in Phase III to determine whether any of these models are considered useful in evaluation.

Phase II Research Question

Phase II asks: *How do evaluators conceptualize audience engagement when creating data viz?*

Phase II Method

This phase will follow the general contours of grounded theory (Corbin & Strauss, 2015; Taylor, et al., 2016) using follow-up interviews to better understand and articulate how practicing evaluators understand the theoretical relationships between the subthemes within the emerging *audience engagement* model. The Claremont Graduate University Institutional Review Board [CGU Protocol #4579] certified this follow-up interview project as exempt.

Phase II Participants

Phase II participants (N = 6) were selected from the 2021 interview participants. They ranged in age from 32 to 47 years old (M = 38.8, SD = 5.5), were primarily female (66.67%), primarily white (66.67%), and all educated with a minimum of a master's degree. Note that demographic details were not requested again for follow-up interviews. Participant age in 2023 was calculated by adding two years to the age provided in 2021. It is possible that participants changed their gender or ethnicity identity between interviews, but this was not captured. The double letter identifiers from Phase I remain, but the years of experience have been increased by two years to differentiate from statements made previously. For example, a quote attributed to VV-15 was made in 2021 through secondary data analysis and one attributed to VV-17 was made in 2023 through follow-up interview.

Theoretical sampling (Corbin & Strauss, 2015) allowed me to select participants I deemed best able to address theory gaps. These data viz and evaluation experts were selected and re-interviewed because they described portions of the audience engagement model in their prior interviews and I felt they would be “information-rich” sources (Patton, 1990) able to elaborate on the perceived subtheme relationships within the emerging *audience engagement* model. The sample size was not determined ahead of time, but it was expected that a minimum of five interviews would be required. Interviews stopped as soon as the conceptual model was described well enough to include it in Phase III.

Phase II Materials

A new semi-structured interview protocol (Appendix E) was designed for this study to explore program evaluators’ conceptualizations of the emerging *audience engagement* model. Specific questions were intended to address the subthemes numbered five through nine of the initial model: (5) *increase satisfaction* (i.e., *What do you think is the role of data viz in increasing audience satisfaction with evaluation?*); (6) *build capacity* (i.e., *How involved are your clients in the data viz process?*); (7) *brand identity* (i.e., *Do you incorporate client brand image and style-guides into your data viz work? If so, how and why?*); (8) *provide artifacts* (i.e., *How do you typically provide data viz to clients?*); and (9) *additional audiences* (i.e., *Do you find your clients using your data viz to share with others? Give me some examples of that.*)

In addition to responding to specific interview questions, participants were shown and asked to respond to the emerging *audience engagement* model in Microsoft PowerPoint via screenshare. In this way, interviewees were asked to “participate in the process” (Patton, 2015) and become “collaborators” (Creswell, 2009). Later interviewees were also provided a three-minute video describing the *efficiency rationale* for data viz. Creation of that video will be described in Phase III. The video was provided via YouTube link in Phase II to serve as a refresher for interviewees on visual efficiency and the information processing model. As the information processing model so clearly underpins the first four themes of the *audience engagement* model, it was more time efficient to send the video for viewing ahead of interviews rather than explaining the concept in each interview.

Phase II Data Analysis

Data in this phase were coded and analyzed in a manner that follows the general contours of grounded theory. Grounded theory (Glaser & Strauss, 1967) is a classic and commonly applied qualitative analysis approach used to develop theory. Given that the aim of the present analysis is to propose a new theory of data viz use, the approach aligns well with the Phase II study aims.

Consistent with general guidelines on conducting a grounded theory analysis, data collection and analysis for this phase were a highly iterative process. I began coding as soon as the first interview was conducted. I made changes to the interview protocol as needed to ensure I obtained the data needed to describe and vet the emerging model. In this study, it was not the mere presence of a theme or the number of times it was mentioned that mattered, but how each coded segment contributed to my understanding of that theme within the context of *audience engagement*. Themes were merged and separated multiple times during analysis. I used the constant comparative approach (Glaser & Strauss, 1967), whereby each newly coded section of data was examined against previously coded data to better understand interviewee conceptualizations. Codes were refined, edited, and omitted to ensure they best fit the data. This method of “joint coding and analysis” (Glaser, 1965) is intended to generate a theory which is "integrated, consistent, plausible, close to the data and in a form which is clear enough to be readily, if only partially, operationalized for testing in quantitative research" (p. 437).

Emphasis was placed on memoing and integrated analysis through constantly comparing new data to existing data rather than merely coding and theming the data. Consistent with a project on data visualization, visual depictions of the emerging model were used extensively in the memoing process. The visual model itself was updated during and following each interview. Although this process was both integrated and iterative, it will be described below in three steps.

Step 1: Create Preliminary Memos from 2021 Interview Data and Identify Gaps

Significant time had passed from the Phase I analysis, so I started by re-immersing myself in the 2021 interview data while focused only on the *audience engagement* model. I started with the coded interviews in MAXQDA 2022 (VERBI Software, 2022) conducted and analyzed in Phase I, the visual model of buckets created through that analysis (Figure 11), and a summary memo describing my then current knowledge of the model. This initial summary memo corresponds to the *audience engagement* model description from Phase I. I reviewed each of the eighteen interviews to compare the knowledge in that interview to the existing model. During that review, the summary memo was updated and/or new theme memo(s) were created as needed to analyze the emerging model of *audience engagement*. I placed special emphasis on the unclear relationships within the model, particularly themes five through nine, which appeared to be the most evaluation specific within the model.

In parallel, I created a list of model questions and potential interview questions that each participant might be able to help address during a follow-up interview. Through this theoretical sampling process, participants were identified as high (n = 7), moderate (n = 3), or low (n = 8) value for follow-up interviews. Their participant memos were color coded accordingly (i.e., green, yellow, or red, respectively) in MaxQDA. This list and associated notes informed the order of the interviews and the starting point of each interview; low value interviewees were not invited to interview. For example, VV-

15 made several statements such as, "… the brain science is that the human brain can only latch on to or remember like a couple of things from a presentation" and my participant memo on that interview included the comment: "Makes heavy use of the information processing model without directly mentioning it. May want this to be a later interview so can share the emerging model."

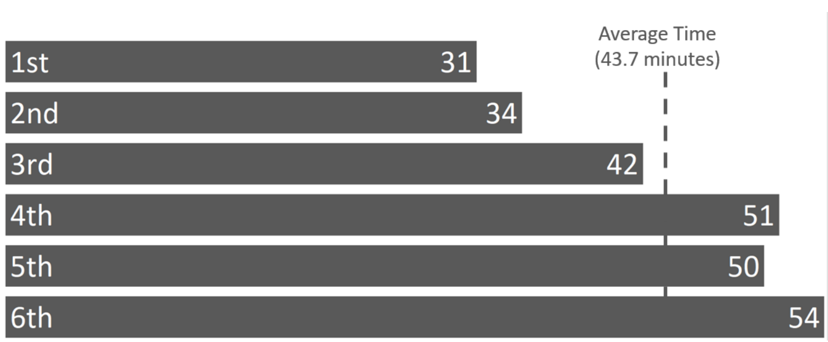
Step 2: Conduct Interviews and Analyze Data, Concurrently and Iteratively

Invitations Sent. In an iterative process, each participant identified for follow-up (*n* = 10) was invited to participate in a follow-up interview via email (Appendix F) and provided with informed consent via Qualtrics (Appendix G). Reminder emails were sent as needed, for a maximum of three email contacts with each person.

Interviews Conducted. Interviews were conducted in the fall of 2023 via Zoom to allow for easy screen sharing of the model as well as convenient transcription of the interview. To allow time for analysis, no more than one interview was held per day. Interview length ranged from 31 to 54 minutes (*M* = 43.7, *SD* = 9.6) with later interviews tending to be longer as there was more discussion on the emerging model that was presented to participants, Figure 13.

Figure 13

Interview Time in Minutes Tended to be Greater in Later Interviews



Consistent with conducting a grounded theory approach, some analysis occurred during the interviews. Following general questions such as, *What does audience engagement mean to you?*, participants were asked questions about the areas of the model needing clarification. During the latter portion of each interview, participants were shown the then current model and asked to clarify specific portions. Using drawing features and tools in Microsoft PowerPoint, I annotated and edited the model in real time while asking questions such as, *Do you think [this] leads to [this]?*, *Can you have [this] without [this]?*, and *How do you think these [two] concepts relate?* This visual collaboration was appropriate in this study because (a) the interviewees are visualization experts and (b) the intended end result is a visual representation of the conceptual model. It was time efficient to present the then current model to them and ask for their direct feedback. As might be expected from this sample, participants had evidence-based opinions on the meaning implied by even the tiniest visual components of the model, such as shapes selected, color choices, and whether an arrow should be bi-directional.

Interviews Transcribed and Analyzed. Starting from the automatically generated transcript provided by Zoom, each interview transcript was reviewed and edited, with the Zoom recording reviewed as needed, to create an accurate transcription. This written transcript was then imported into MaxQDA and used to create a participant memo summarizing the key contributions of each participant and to note any gaps remaining that the particular participant might be able to clarify in a follow-up email or additional interview. Theme memos were also updated at this time. This process was completed immediately following the interview.

Step 3: Conceptual Model Revised After Each Interview

In addition to the analysis that was taking place during and immediately following each interview, analysis and revision of the written and visual conceptual model took place in the days following each interview. Coding and memoing continued in MaxQDA for the written model, but the Rich Text Format (RTF) document file native to MaxQDA does not have sufficient graphics capability to create the rich visual representation of the emerging model that I desired for this project. Therefore, I continued developing the visual model in Microsoft PowerPoint following each interview. This process involved translating the written understanding into a visual representation. For instance, Figure 14 shows the emerging visual model after the second interview and Figure 15 shows the evolution of the conceptual model after the fifth interview.

Figure 14

Example of Emerging Visual Model (Draft) After Two Interviews

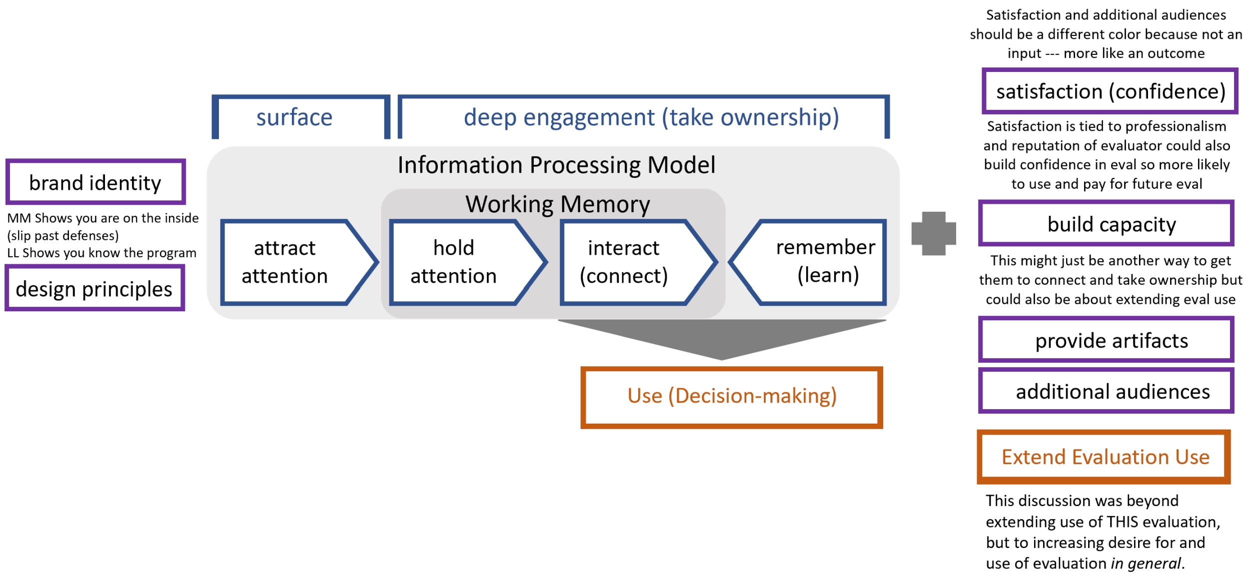
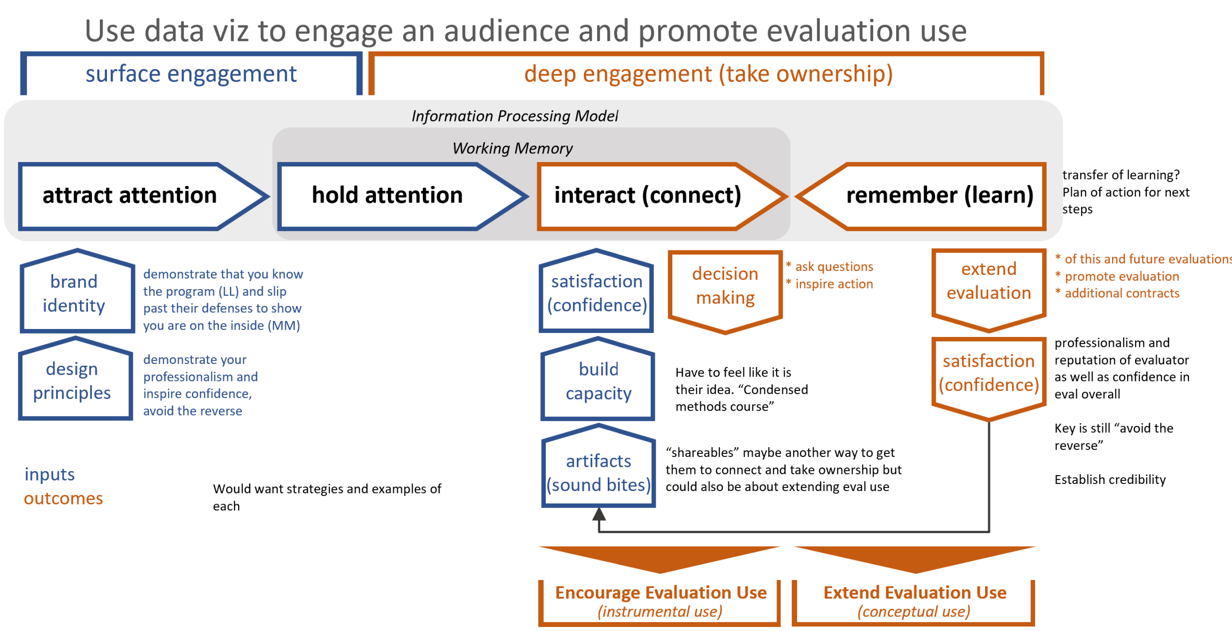


Figure 15

Example of Emerging Visual Model (Draft) After Five Interviews



This heavy emphasis on visual memoing was appropriate for this project on conceptual models in data viz. Verdinelli and Scagnoli (2013) emphasize the role of creating visual representations of emerging theories as "an intrinsic and essential step in theory building" (p. 14) and, consistent with my interest in demonstrating the role of data viz in analyzing qualitative data, I wanted to take this opportunity to demonstrate this method of articulating a mental model. This intentional emphasis on using data viz in my analytical process was also motivated by this quote from Corbin and Strauss (2015) of Miles and Huberman (1994) on the benefit of creating a visual conceptual framework:

> Conceptual frameworks are best done graphically, rather than in text. Having to get the entire framework on a single page obliges you to specify the bins that hold the discrete phenomena, to map likely relationships, to divide the variables that are conceptually or functionally distinct, and to work with all the information at once (p. 123).

After an updated visual version of the model had been created and all memos were updated, only then was the next interview invitation sent. As soon as the model was described well enough to present in Phase III, no further interviews were requested.

Phase II: Findings and Discussion

Insights from Phase I concerning audience engagement and findings from this Phase II are merged with discussion in this section. Consistent with the qualitative methods used to uncover these findings, these data are highly interconnected and embedded within their context. As these findings emerged through my interpretation of the data, I hope that discussing my interpretation of these findings in the same section will provide a more holistic analysis and be easier to read. To ensure clarity, the findings and discussion will be presented thematically – in the categories of the final model. This is intended to be an evidence-based interpretation of the data. For transparency, I will endeavor to clearly differentiate between the presentation of findings (quotes from participants), presentation of examples from literature that corroborate the findings, and the discussion (my interpretation and synthesis between interviews).

Naming the Model: Audience or Stakeholder

The term audience is used throughout the description of this model even though the term stakeholder is more commonly used in evaluation. The rationale for using the term audience is two-fold. First, the original interviews analyzed in Phase I often used the word audience instead of stakeholder and so this terminology and conceptualization was specifically explored in Phase II follow-up interviews. Second, this research coincides with evolving discussions within the evaluation community regarding concerns over the word stakeholder.

The Term Stakeholder is More Common in Evaluation Literature Than the Word Audience

A literature search conducted in December 2023 of the three journals most prevalent in evaluation, *The American Journal of Evaluation* (AJE), *New Directions for Evaluation* (NDE), and *Evaluation and Program Planning* (EPP), supports that *stakeholder engagement* is a much more common term than *audience engagement*. This simple search method only considered the appearance of the word(s) in any portion of the article text over the publication time span of each journal.

In that search, there was only one result for the specific term "audience engagement" in each of the three journals reviewed. One AJE article referenced two articles from outside of evaluation to support "including digital stories in reporting increases the potential to effectively use findings through deeper audience engagement" (Abbato, 2023); one NDE article was an editor's note referring to audience discussion at a conference (Cousins, 2007) and is not applicable to this study; and one EPP article discussed fostering audience engagement through participation in a data collection activity using post-it notes displayed visually on a wall (Long et. al., 2022). Comparatively, there were significantly more returns for "stakeholder engagement" in each journal and the word stakeholder is used in more articles than the word audience, see Table 5. Other combinations such as "engage stakeholder(s)" do return additional articles which were not analyzed here; this is not intended to be a literature review, merely a demonstration of the prevalence of the word stakeholder over audience in evaluation literature.

Table 5

Frequency of Terms in Eval Literature by Number of Articles Using the Term

Search Term	AJE (1981-2023)	NDE (1997-2023)	EPP (1978-2023)
"audience engagement"	1	1	1
"stakeholder engagement"	38	41	65
audience	356	346	715
stakeholder	1,030	734	1,344

Note. This search considers the appearance of the word anywhere in the title or body of the article but not references.

Audience was Used by Phase I Interviewees and was Specifically Addressed in Phase II

A review of word choices used in Phase I shows that 44% (*n* = 8) of interviewed evaluators used the word *audience* exclusively and did not use the word *stakeholder* in their interviews. An equal number (*n* = 8) used both words. Only one interviewee used the word *stakeholder* exclusively and one other did not use either word, using the word "client" instead. This emphasis on *audience* even though *stakeholder* is more common in evaluation, prompted adding the question: *Do you prefer the term audience or stakeholder? And Why?* to Phase II follow-up interviews (Appendix E) so that the distinction could be addressed directly.

Of the six follow-up interviews, three had used audience exclusively in their first interview and three had used both audience and stakeholder in their Phase I interview. In Phase II follow-up interviews, participants indicated that the term *audience* is "broad" [MM-9] and "passive" [YY-7; ZZ-12], while *stakeholder* is more "technical or exclusionary" [MM-9], but also connotes action and engagement. The passivity of *audience* was described as: "sitting in a movie theater and watching a movie. There's not the interaction between what's on the screen…and the audience" [YY-7].

Audience appeared to be used as an umbrella term that can include *stakeholders*. One participant differentiated between internal *staff* and referred to external partners as *stakeholders*, but then combined these separate conceptualizations into one word saying, “you know they're different *audiences*, and I care about all of them when I create visuals” [KK-12].

Some Evaluators Have Expressed Concerns with the Word Stakeholder

This research is being conducted during an unfolding discussion within the evaluation community regarding concerns over use of the word *stakeholder*. This discussion is occurring in private dialogs, blog posts, and conference sessions but has not yet been presented in peer reviewed evaluation literature. The word does have economic origins with the earliest reference in the *Oxford English Dictionary* from 1709 referring to the person or organization holding funds deposited by others on a bet or financial transaction. By 1821, the meaning had become “a person, company, etc., with a concern or (esp. financial) interest in ensuring the success of an organization, business, system, etc.” (Oxford English Dictionary) and is similar to how the word is commonly used in business today (Clayton, 2014). In business, the broad term *stakeholder* can be applied to employees, community members, etc. as anyone who has an interest in the activities of the company and is meant to differentiate from the legal term stockholder who holds a clearly defined financial stake in the company (Kujala et. al., 2022).

In evaluation discussions regarding the word *stakeholder*, concerns over the word relate to its “mercenary connotation,” the imbalances suggested by one group holding the money or power of the other, and concerns of Indigenous Peoples who are better referred to as “rights and title holders” (MacDonald & McLees, 2021). However, the concern that the word is a “catchall phrase” (MacDonald & McLees, 2021) and too nebulous to be useful, is precisely why it is valuable; it is intended to encompass anyone who may have an interest in the evaluation.

Controversy over the word and frustration that no better word has emerged were apparent in Phase II follow-up interviews. Only one interviewee was unaware of the controversy surrounding the word stakeholder. There was a consensus that the word is “not ideal” but that they “haven’t found another word or phrase that I like better” [ZZ-12]. This participant quote exemplifies the frustration inherent in identifying alternative terminology:

> I'm personally trying to move away from the term stakeholder. … There's a negative connotation … So, I'm thinking clients, program implementers, program managers, funders…The term just feels a lot broader to me, like everybody vs. [long pause] I mean, I like the term *stakeholders* in that like it is literally those who have a stake in something, and I wish there were a better term.

The Term Stakeholder Would Also be Applicable to This Model

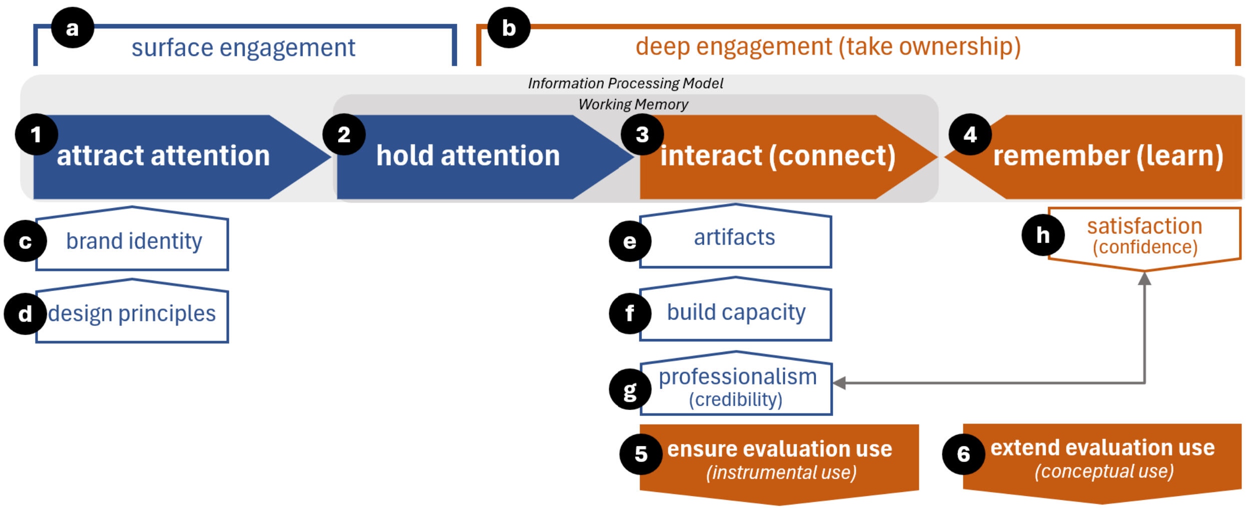
Although I will use the word *audience* in this model, the reader can substitute the word *stakeholder* if they choose. While *stakeholder* is still the term commonly used in evaluation, *audience* is appropriate for this model because it was considered more broad, less technical, and less controversial.

A model of Data Viz for Audience Engagement and Evaluation Use

The model of using data viz for *audience engagement* (*AE*) and to ensure evaluation use that emerged in Phase I and is described in this phase is depicted in Figure 16. A brief overview of the model is presented and then each portion is described in detail. The model can be considered as two intertwined layers. The core of the model is the information processing model of memory presented in four steps from an evaluation point of view and two steps representing the ultimate expected goals of the model. Intertwining with that core model are evaluation specific conceptualizations of activities presumed to engage the audience and ensure evaluation use.

Figure 16

Model AE: Using Data Viz to Engage an Audience and Ensure Evaluation Use



Note. Black circle identifiers are labeled for this discussion only. They are not part of the final model and not included in the model shared with participants in Phase III.

In this model, the first level of *audience engagement* can be thought of as (a) *surface level engagement* wherein the goal is simply to cut through the ever-present visual noise of our busy world and (1) *attract* or grab the audience's attention. The suggested methods to (1) *attract* this attention are good use of color, motion, (d) *data viz design principles*, and intentional use of (c) *brand identity*. Successful use of (d) *design principles* and (c) *brand identity* can also demonstrate evaluator (g) *professionalism*, inspire (h) *confidence*, and (2) *hold* the audience's attention.

Wise use of (d) *design principles* and (c) *brand identity* may help (2) *hold* audience attention long enough to move the audience into a stage of (b) *deep engagement* during which they are (3) *interacting and connecting* with the data being shared with them. In (b) *deep engagement*, the audience is taking (b) *ownership* of the data, asking more questions, and making decisions based on the data. The hoped for result of this (3) *interaction and connection* with the data is (5) *instrumental evaluation use* whereby the audience makes decisions based on the data presented in the visuals and is inspired to action.

Interviewees suggested a number of opportunities for an evaluator to encourage (3) *interaction and connection*. An evaluator may plan data viz so that the audience has (e) *artifacts*, soundbites, or shareables to use for secondary purposes and to share with additional audiences. Building the audience's (f) *capacity* to interpret, use, and create data viz is another technique that evaluators can use to keep an audience (b) *deeply engaged* in the data to the point that they feel *connection* and *ownership*. The (g) *professionalism* and skill with which the evaluator creates and manages this (3) *interaction and connection* with the data viz also contributes to the success of this model.

Another aspect of (b) *deep engagement* is for the audience to (3) *connect* with the data in such a way that it becomes a part of them – that they deeply understand, learn, and (4) *remember* the information they connected with. The ultimate goal of this portion of the model is to (6) *extend evaluation use*. Extending evaluation use can mean *conceptual use* whereby the audience has changed their knowledge or understanding based on the data viz and applies that knowledge later, maybe even to another program. Or extending evaluation use can be thought of as promoting the specific evaluator or evaluation in general. This could mean that the audience purchases additional evaluation contracts from that evaluator or that they purchase future evaluation contracts from other evaluators. Creating satisfaction and confidence in the data viz, the data, the evaluator, and the evaluation as a whole was discussed as a way to elevate the field of evaluation (and thus encourage use of evaluation as a whole) and as a way to promote a particular evaluation firm.

In this model, audience (h) *satisfaction* and *confidence* are closely related to the (g) *professionalism* and *credibility* that the evaluator brings to the model. In a feedback loop, the audience is more (h) *satisfied and confident* because the evaluator demonstrates (g) *professionalism* and the evaluator is perceived as more *professional and credible* because the audience is *confident and satisfied*. As with (6) *extending evaluation use*, this confidence is not limited to the particular evaluator or the specific evaluation – the confidence extends to the profession as a whole.

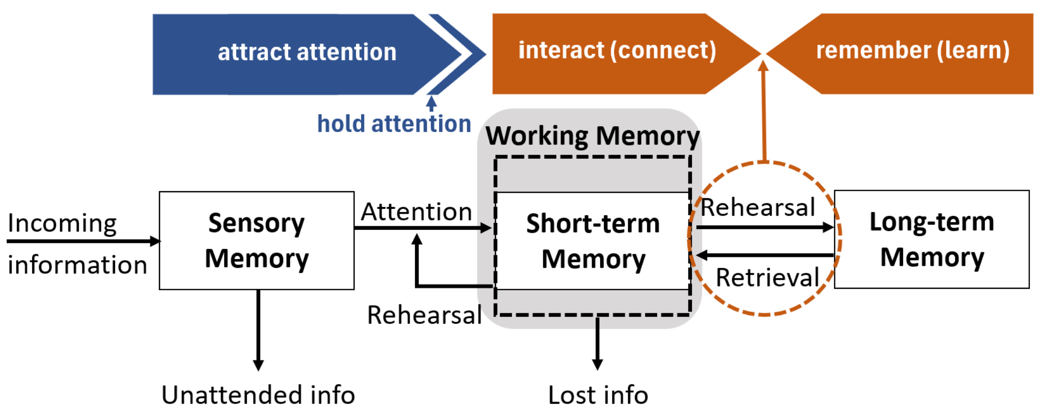
The Information Processing Model in Evaluation Context

The four color-filled arrow boxes in the model labeled 1 through 4 (see Figure 16) were clearly identifiable in Phase I findings. Their naming and perceived order were not changed through Phase II follow-up interviews, but additional information concerning the audience's anticipated travel through these stages and the relationships between the stages was gleaned. Figure 16-a

In this portion of the model, data visualization is used to (1) *attract attention* and then (2) *hold that attention* long enough that the audience can (3) *connect* with the information presented. Connecting with the information was also discussed in terms of taking ownership of the data and interacting with it for deep understanding. There is clear overlap between these conceptualizations and the Information Processing Model of Memory depicted by Huang and colleagues (2009) when describing cognitive load theory. That model is presented in Figure 1 earlier in this paper and is presented here as Figure 17 with the audience engagement model superimposed over it. In this way, steps one through four of the *audience engagement* model can be thought of as the information processing model from an evaluator point of view.

Figure 17

The Information Processing Model of Memory Combined with the Audience Engagement Model



Note. Black and white model is a recreation of the Huang et al. (2009, p. 3) visual. The gray, orange, and blue boxes are portions of the audience engagement model stretched to show alignment to that model.

Stage (1): Attract Attention. An initial or *surface level of engagement* was discussed as a way to catch the eye and get attention from the audience in a world filled with information vying for that attention. *Attracting attention* relies on the principles established earlier in this paper: humans are vision-based animals and prioritize color, images, and other visual data over text (Few, 2012; Hegarty, 2011; Ware, 2021). This presumption that a visual image will attract more attention is supported by the well-known priority given to texts containing photos over those containing only text in social media (Li & Xie, 2020) and marketing (Pieters & Wedel, 2004). This uncontested assumption that humans prioritize images over text is also consistent with Evergreen's emphasis on visual design in evaluation reporting (2011; 2013; 2017; 2018). In terms of the Information Processing Model of Memory (Huang et al., 2009), this is pre-attentive processing – using color, movement, and Gestalt principles – occurring in sensory memory to sort incoming information into either discarded "unattended info" or move it into the "attention" stage.

Stage (2): Hold Attention. On its own, *attracting attention* is not considered enough to move an audience into *deep engagement*. Interviewees felt that once the data is attended to and attention is gained, something additional is required to hold that attention. There was no clear demarcation discernable between *attract* and *hold*, nor between *hold* and *interact*, but there is a presumption that an audience does not move directly from *attract attention* to *interaction*.

Interviewees emphasized the need to *hold* audience attention with the data viz, but implied that *holding attention* is not important on its own – it is not an end goal of data viz – it is only a transitional step in moving the audience to the desired next step, *interact (connect)*. Rather than identifying specific activities that *hold attention*, discussions suggested that the activities that can *attract attention* (i.e., design principles and brand identity) and the activities associated with *interact (connect)* also play a role in *holding attention*.

My explanation for the ill-defined nature of the *hold attention* step of this model is that this step cannot be considered on its own. It is a concept that bridges a gap between *attract attention* and *interact* and serves as a reminder to evaluators that *attracting attention* does not automatically lead to *interaction* – something more is required. I suggest that evaluators think of *hold attention* as overlapping the very end of *attract attention* and the very beginning of *interaction*. Focusing on successful activities in this area of the model can help to move an audience from *attract* to *interact*.

Stage (3): Interact (Connect). The reason to *attract* and *hold attention* is to move the audience into the third step of the model wherein they *interact and connect* with the data underlying the data viz in a deep way. The conceptualization of *connection* is both short and long-term. Examples of short-term activities or signals showing that *connection* is taking place are the audience asking additional questions of the data, the evaluator, or each other in a meaningful way. Participants suggested that engagement can also cover a very long period of time and considered long-term feedback on the data or evaluation questions addressed through the data viz as continued engagement: "If they come back with suggestions like six months after you've [provided a visual and say] 'I'd love to see this idea presented in the same way, or… to be able to look at the data in a different way, or to be able to toggle between X view and Y view'" [YY-7]. The importance of this stage can be discussed around three different themes: ownership, the "ah ha" moment, and instrumental use.

The Interact (Connect) Stage was Closely Associated with Ownership. The word "ownership" was used in multiple interviews from Phase I and II to describe the deep level of connection that was desired in this stage of the model. The phrase (b) *take ownership* was added into the model as a component of *deep engagement* spanning both (3) *interact (connect) and* (4) *remember (learn)* as ownership was associated with both these stages. This sense of *ownership* instilled through activities associated with data viz is believed to extend to the evaluation as a whole, "…it helps them feel a little sense of ownership over their own work, or over the evaluation that you've provided them" [LL-6].

A number of evaluator actions are believed to promote this sense of *ownership* through *interaction (connection)* with the data viz including *providing artifacts*, *building capacity*, and *demonstrating professionalism*. Each of these themes is discussed later.

Connection is an Internal Process Reminiscent of the "ah ha" Moment. The value of the "ah ha" or "wow" moment in creating engagement and forming memories is not evaluation specific nor data viz specific. While traditional problem-solving relies on working memory, executive processes, and long-term memory; non-traditional problem-solving (which I associated with the "ah ha" moment of the *data* → *insight* model in my literature review) has the audience making the mental "leap" themselves and this leap or jump to a solution is accompanied by neural activity associated with the dopaminergic midbrain's reward process (Chang et al., 2009; Tik et al., 2018). Without necessarily knowing these underlying processes, interviewees discussed similar activities in the context of *audience engagement*. When talking about locking a story into memory, VV-12 indicated that there is one single soundbite or take-away that "I'm intentionally kind of building up to – that when it hits – that's the memorable thing, and oftentimes I will leave blank space and silence there to just give people like time with it." These "high leverage situations" [VV-12] show a clear link with the "ah ha" moment of the *data* → *insight* model and the dopamine release related to locking connections into memory.

This type of deep engagement is also suggested by Kirk (2019) as a component of a third category, *exhibitory*, within the *explain* ←→ *explore* model in which the data viz designer asks the audience to fill in the blank. An example of designing a viz in a way that invites an audience to make the cognitive leap themselves is provided by Kirk and used here with permission. Figure 18 visualizes the one hundred highest paid athletes in 2018.

Figure 18

Highest Paid Athletes in 2018 Data Viz by Andy Kirk

It would take far less room to simply state that the one hundred highest paid athletes in 2018 were all male, but both impact and memorability could be lost. This example also serves to demonstrate the close relationship between model steps (3) *interact (connect)* and (4) *remember (learn)* and explains why the directional arrows of these two steps point toward each other in a reinforcing relationship.

The Ultimate Goal of Interact (Connect) is Instrumental Use. All participants in Phase II follow-up interviews suggested a relationship between getting an audience to *connect* with the data via the data viz and then translating that *connection* into *instrumental use*. The definition of *instrumental use* differed by audience with statements like wanting program staff to "make a decision based on it" [KK-

12] or "compel them to some action" [MM-9], or "if it's a high network donor like, I want them to write a check, right?" [VV-17].

Decision-making was the desired outcome most often discussed at this stage. However, there was no consensus on how to assist an audience toward decision-making. For example, some participants stressed the importance of clearly stating a best course of action, while another discussed the importance of providing enough "scaffolding" for the audience to make a decision on their own, without being directive about the action to take, as "it's gotta feel like their idea to be successful, right?" (ZZ-12). Yet another participant discussed a case of *misuse* in which a client used the presence of a prioritized list as a political tool to prevent funding of any items on the list because there was insufficient funding for the first item on the list.

There is Overlap Between Interact (Connect) and Remember (Learn). What was clear from discussion in both Phase I and Phase II is that there is overlap, or even possibly a feedback loop, between steps (3) *interact (connect)* and (4) *remember (learn)* as portrayed by the reference arrows in the model. An inability to clearly separate these two stages is acceptable considering that the Information Processing Model of Memory (Huang et al., 2009) also clearly links these two stages through the concepts of "rehearsal" and "retrieval" (see Figure 17).

Relatedly, there is a great deal of conceptual overlap between Stage (5) *ensure use* and Stage (6) *extend use*. For conceptual clarity, this model portrays (3) *interact (connect)* as most directly related to *ensuring use* and (4) *remember (learn)* as most directly related to *extending use*, but discussion below will demonstrate the strong relationships between all four concepts in this model.

Stage (4): Remember (Learn). The idea that evaluators want the audience to *remember* key points of the data viz and *learn* from the data viz is straight-forward. The activities that were suggested to assist with *remember (learn)* can just as easily be associated with *interact (connect)*, so the model is drawn with the activities grouped together. The strong relationship between (4) *remember (learn)* and

(3) *interact (connect)* may make it more difficult for an evaluator to separate these concepts, but separation of these stages may not be necessary in practice. Many of the *connection* activities discussed directly related *connection* to locking that piece of information into *memory*, which fits well within a literature backed explanation of memory mechanisms. These technical explanations are previously described in the *data* → *insight* model and their prevalence here may suggest that evaluators are embedding this *insight* conceptualization of data viz into the *audience engagement* model without directly identifying it.

Surface Engagement Is Different from Deep Engagement. Re-analysis of the original interviews in preparing for the follow-up interviews made clear that there was an over-arching distinction between (a) *surface level engagement* and (b) *deep engagement*. This distinction became even more clear as Phase II follow-up interviews proceeded and these demarcations were added to the model.

Attracting attention was more closely associated with *surface engagement* with terms like: “I've got them. They're hooked. … I've got their undivided attention” [VV-17] and “it really sucks you in. … and you're like totally intrigued” [LL-6]. Interviewees clarified that there is another level of “participation” or *deep engagement* after *surface engagement*, “… that next step which is like they're coming up, they're doing data walks, or it's a live QA, or it's a lot of back and forth, or they're doing stuff like building their own visualizations or getting deep into the data” [VV-17] and it is into this category that both the *interact (connect)* and *remember (learn)* steps of the model are contained.

Deep engagement was characterized by asking questions, taking *ownership*, and moving beyond the data viz into the information and meaning behind the data. It was suggested that the audience is not deeply engaged if they are still critiquing and discuss the data viz itself: “if they're still critiquing your colors, they probably haven't fully engaged” [YY-7]. In other words, if the data viz has done its job well, it is aiding understanding and not getting in the way. In this sense, the data viz may no longer be the center of attention during this important part of the model.

Just as there is no clear demarcation between *hold attention* and *interact*, the demarcation between *surface engagement* and *deep engagement* is unclear. Early versions of the model depicted *surface engagement* as equal to *attract attention*, but one participant suggested that the boxes be drawn to overlap *hold attention* to represent that there is not a clear demarcation between *surface* and *deep engagement*.

It should be clear that the first four steps of the model are not specific to evaluation. I characterized them as the Information Processing Model of Working Memory from an evaluation point of view and that conceptualization became represented in the final visual design of the model. While sharing the emerging model with interviewees, it became apparent that – whether they were familiar with the information processing model or not – this model resonated with them and felt applicable to their evaluation work. I hope that articulating these steps as conceptualized by evaluators offers helpful insights into how evaluators believe these steps can be achieved and helpful hints for other evaluators to incorporate into their work. The next themes to discuss are the evaluation specific portions of the model, themes (c) through (h) as well as stages (5) and (6).

Evaluation Specific model of audience engagement with heavy emphasis on use

Intertwined with the Information Processing Model of Memory presented from an evaluation point of view described above are evaluation specific conceptualizations of activities presumed to engage the audience and ensure evaluation use.

Design Principles. Use of *design principles* was not identified as a separate sub-theme in Phase I, but through Phase II interview discussions, it became apparent that data viz design principles are not simply a component of *attracting attention*. Wise use of *design principles* is highly inter-related with decisions on how to incorporate *brand identity*, with the audience's perception of evaluator *professionalism* and competence, and with the *interact (connect)* portion of the model. Using *design* *principles* wisely is most obviously associated with *attracting attention* and so is placed at the beginning of the model, but these considerations are important throughout.

Brand Identity. Intentional use of *brand identity* is another important consideration at this early stage. Using intentional branding over a randomly selected design scheme is also believed to *hold attention* long enough to create *interaction* and was associated with *satisfaction* and *professionalism*, as well. Every interview in Phase I and follow-up interview in Phase II discussed intentionality around using *brand identity* in data viz.

Decisions around incorporating brand identity were complex and rife with challenges. Many of these challenges, such as compromising best practices in data viz design to match client logo colors [GG-10], are detailed previously by Douville and colleagues (in press) and are discussed in Phase I. In Phase II, interviewees were specifically asked about how and why they make *brand identity* decisions and how they believe it relates to *audience engagement* (Appendix E). Once the emphasis on "challenges" was removed from the conversation, interviewees discussed their thought processes around whether to use the client's brand identity or the evaluator's own brand identity. While using the client's branding was discussed as a way to "get [the data] past their defenses" [MM-9] under the assumption that clients have more confidence and trust in an internal document, the desire to be perceived as internal needs to be carefully weighed against the perception of "objectivity" [YY-7; ZZ-12] that some clients want to achieve by hiring an external evaluation firm.

The most common scenario discussed is evaluators using their client's or donor's *brand identity* because it *attracts attention* and helps the audience feel a sense of *ownership* over the data and the evaluation results [LL-6; MM-9]. It was also suggested that using client branding demonstrates that you know the program, may help overcome resistance because the document appears to be internal, or may make it "easier for them to pass along internally" [MM-9].

Intentionally using the branding of their own evaluation firms in data viz was only discussed by the two interviewees who work with prestigious and internationally recognizable evaluation firms. In both interviews, the evaluators occasionally use their firm's brand identity because they believe funders and stakeholders want the prestigious logo placed near their own logo in the suggestion of a partnership. They suggested that the *credibility* associated with their *brand identity* can be used to *hold attention*. Both participants also discussed using their client's branding in other circumstances, but neither discussed blending the brand identities.

All participants discussed these *brand identity* decisions as if they were dichotomous, which does not fit with the complexity possible within branding decisions. For instance, Evergreen and Sabarre (2019) show that *brand identity* extends far beyond color choice and logo to include fonts, layout styles, and even aesthetic choices such as whether visuals appear hand-drawn, have rounded edges, etc. The emphasis in this project was on how evaluators feel *brand identity* affects *audience engagement*, so a thorough conceptualization of what is meant by *brand identity* was not sought. A better understanding of this topic is an opportunity for further research.

Interestingly, although every participant intentionally considers decisions around branding and all participants suggested ways that branding decisions support their data viz goals, many hinted that they do not have any evidence to support their beliefs about whether and how *brand identity* decisions affect *audience engagement* with statements such as, "I don't have feedback that that happens. I just think that it did." [MM-9] and "I think it's some legitimacy to it, I think." [YY-7]. This is also an opportunity for further research.

Artifacts. Phase I interviews discussed the importance of providing data viz materials that could be used independently from the evaluation process and report. I referred to these graphics, presentation slides, hand-outs, etc. as *artifacts* and considered their creation an important theme of the emerging *audience engagement* model. Through Phase II discussions, it became apparent that the *artifacts* themselves were not a major theme of the model; *artifacts* are just one more opportunity to deeply engage clients, give them a sense of *ownership*, *encourage use*, and *extend evaluation use*.

When asked directly about the role of *artifacts*, Phase II participants appeared to downplay the significance that I had felt in reviewing transcripts of earlier discussions. Now, they emphasized that the artifact didn't necessarily have to be a tangible piece of data viz – it could be a soundbite [VV-17] or simply the one bit of information that you want the audience to remember and discuss later [KK-12; MM-9; YY-12]. I believe that I misinterpreted the importance of the role of *artifacts* in Phase I because the directed content analysis (Hsieh & Shannon, 2005) method used emphasized the number of occurrences of the theme as saturation, whereas applying a grounded theory approach in Phase II emphasized the relationships between the themes (Corbin & Strauss, 2015; Hennink et al., 2017; Hennink & Kaiser, 2022; Kerr et al., 2010). In this analysis, it became apparent that *artifacts* are mentioned frequently because they are useful at many points throughout the model, not because it is important as a stand-alone theme.

Phase I closely associated *artifacts* with *additional audiences* because they were often shared with these additional or distal audiences. Phase II clarified that the evaluators interviewed do create stand-alone data viz pieces for their primary audience with the intent to have that viz shared with distal audiences, but they stressed the importance of understanding and controlling for the additional audiences from the beginning so that the clients – as the owners of the data – will come back to the evaluator before repurposing the data viz [KK-12]. They recommended against providing data viz that can be used in a multitude of ways. Rather, they stressed engaging stakeholders throughout the process, "so that you can kinda suss out beforehand, not after the fact" how they will use the materials provided [YY-7] and then working with the client to ensure they understand the data viz and underlying message well enough to share it themselves, believing it will "hit harder" if in their own words [KK-12]. This tied *artifacts* strongly to *building capacity* to interpret and reuse those *artifacts* and also to engaging clients in additional evaluation activities that could lead to additional contracts. One evaluator discussed that their firm has become known as a specialist in creating a particular complex type of data viz that clients regularly share with distal audiences and that clients will often return to them to have this same analysis done on a different data set and referred to this as “establishing us as thought leaders and experts … and at the same time we're giving them incentive to come back to us for additional formal work” [VV-17].

Artifacts are presented in the model (Figure 17) as a way to encourage *interact (connect)*, but it is important to keep in mind that they are useful throughout the model. Providing artifacts is not a single activity – it is done collaboratively with an engaged audience and is believed to eventually contribute to Stage (6) *extending evaluation use*, as will be discussed later.

Additional Audiences Subtheme was Delimited from the Model. Very early in Phase II model development, see Figure 14, *artifacts*, *building capacity*, *satisfaction*, and *additional audiences* were simply added to the end of the model and associated with *extending use*. Through discussions attempting to better understand their perceived roles, it became clear that subtheme categories such as *artifacts* and *building capacity* are useful at various stages within the model and were moved into the model based on those discussions. However, participants only discussed the role of *additional audiences* in the context of *extending evaluation use*; suggesting that additional audiences would lead to additional use. For that reason, the theme of *additional audiences* was delimited (Glaser, 1965) or removed from the model and is only discussed within the context of *extending evaluation use*.

Build Capacity. Similar to the role of *artifacts*, *building capacity* was often referenced in Phase I, but there was little clarity on how it is expected to engage an audience. Phase II discussions clarified that building capacity is also a less important subtheme in the *audience engagement* model than assumed from Phase I. As with *artifacts*, *building capacity* is ultimately perceived as another way to give clients the tools to *extend the evaluation*. It was moved to the middle of the model through discussion because, as stated most clearly by VV-12:

> *capacity* and provide *artifacts* are part of *interact (connect)*, but then they extend past that middle part of the model, because then – in theory – other stuff can happen after the presentation or interaction with the data visualization. If things work as intended, right? …it has kind of that lasting effect or that post presentation effect…

Although it is visually associated with the *interact (connect)* stage, *building capacity* has implications throughout the model. It was suggested that clients want to learn and grow and so opportunities to build capacity increase their excitement, sense of *ownership*, and *satisfaction* with the evaluation and is believed to eventually contribute to Stage (6) *extending evaluation use*, as will be discussed later.

Satisfaction (Confidence) and Professionalism (Credibility). The importance of *client satisfaction* was noted in Phase I, but this theme was greatly expanded upon and divided into two interconnected themes through Phase II follow-up interviews. *Satisfaction and confidence* in the data viz, the evaluator, the specific evaluation, and the field of evaluation as a whole was seen as inextricably linked with the *professionalism and credibility* of the evaluator. Good use of the model subthemes (*design principles*, *brand identity*, *artifacts*, and *capacity building*) is believed to make the audience more *satisfied* while also making the evaluator appear more *professional and credible*. An arrow suggesting a feedback loop was added to the model after several participants suggested that a more *professional and credible* evaluator makes the client more *satisfied and confident*.

Although these two interconnected themes were seen as very important, even critical to *ensuring use*, there were very few specific examples provided on how to increase *satisfaction* or appear more *professional* using data viz. Consistent with the Phase I suggestion that using data viz poorly can harm use and credibility, interviewees emphasized avoiding being unprofessional as opposed to explaining how to appear professional. When pushed to give an example and explain why *satisfaction* and *professionalism* matter to the model – every participant readily provided examples of what not to do. Discussion on how to increase *satisfaction* invariably de-emphasized data viz and focused on the evaluation as a whole with statements like, "I think what's most important is that they're satisfied with the process" [YY-17].

More broadly, the evaluators interviewed showed a clear interest in pleasing clients and linked clients being happy or satisfied with their eventual use of the evaluation. This was presented as a series of "if, then" statements by LL-6 who connected data viz engagement and enjoyment (satisfaction) directly with evaluation engagement and use:

> Yeah, if they engage with the report, then they've engaged with the broader evaluation. Then, they are more likely to use the evaluation for process use, conceptual use, instrumental use, whatever type of use. If they're engaged by the evaluation process, of which data viz is one way we can help engage them in the overall process, then they might be more likely to use evaluation in the future or to commission an evaluation in the future, thereby increasing their evaluation capacity, their evaluation use, and so on. So, [data viz] feels minor. But like, what can we do to make the evaluation process something that they actually enjoy being part of? And I think data viz can help with that.

Ensuring and Extending Evaluation Use. While not included in the Phase I model, the importance of the themes (5) *ensure evaluation use* and (6) *extend evaluation use* became very apparent through Phase II follow-up interviews. When asked about their "ultimate reason" for engaging an audience, every single interviewee tied engagement to evaluation *use*. However, even this small sample of six evaluators had extremely varied interpretations of *use* that were situation and audience specific. And, while they clearly prioritized *instrumental use*, they were optimistic that even the most vague and distal *use* was still evaluation *use*.

Conceptualization of Use was Situation and Audience Specific. Consistent with Phase I findings, participants described a wide range of activities and outcomes that they considered evaluation *use*. These ranged from the expected examples aligned with decision-making and *instrumental use* to highly audience specific conceptualizations of use such as “for policy makers, I want it to be memorable … [for fellow researchers] it should inform their work … [I want] donors to write a check” [VV-17], and for “board members the ultimate goal is … for them to think about our place within the sector broadly” [MM-9].

Five of six interviewees specifically used the phrase “it depends” when describing the *use* intended with the data viz by saying:

- “and even then it's going to depend on who I present to” [LL-6],
- “I think it depends on who the audience is” [MM-9],
- “I think it depends a little bit on the audience” [VV-17],
- “it really depends on the project, the partners, the relationships” [YY-7], and
- “it depends on the purpose of the product” [ZZ-12].

The only person who did not use “it depends” works with an eval firm that has established a “stakeholder analysis tool” that clearly identifies the audience prior to even contracting the evaluation.

Evaluators Prioritize Instrumental Use. Consistent with Phase I, there was still a strong emphasis on *instrumental use* over *conceptual use*. Regardless of the many ways that participants conceptualized *use*, they still prioritized decision-making and *instrumental use*. Even when specifically discussing the role of *remember (learn)* and advocating distal use such as “transfer of learning” [ZZ-12], “inform their work” [VV-17], “awareness of what the organization does” [MM-9], or “more likely to commission an evaluation in the future” [LL-7], participants would invariably make a reference to wanting decision-making and *instrumental use*.

Therefore, the *audience engagement* model portrays *instrumental use* as part of stage (5) *ensuring evaluation use* and places it in the earlier *interact (connect)* portion of the model because direct, *instrumental use*, appeared most desirable to participants. However, it is important to keep in mind that *instrumental use* was the most desired type of use, but it was not considered the most common type of *use* to occur. There were many more examples of *use* presented in the next stage of the model (6) *extend evaluation use*.

A Broad Theme of Extending Evaluation Use (Conceptual Use). The broad concept of (6) *extending evaluation use* became apparent early in Phase II analysis, but I struggled with how to portray it in the model. Early on, it was a category tacked onto the end of the model (see Figure 14) with nearly every other evaluation specific sub-theme under it. As interviews progressed, I began gaining a better understanding of the extremely broad conceptualization of *use* that I am referring to in this project as *extending evaluation use*. It is important to note that no participants used the term "extend" in the context of *extending evaluation use* before I used the term with them, but those that saw the category in the visual model all responded positively to the idea with responses such as:

> I would say, that's not surprising, because … you're often not doing the same thing with the same people over again. So, of course, it's going to extend it beyond that specific project and into a deeper evaluation context, or love of evaluation, cheerleader of evaluation, whichever term you want to use. So yeah, that makes a ton of sense. [KK-12]

As I came to understood this as an ultimate goal of engaging an audience, themes such as *building capacity* and *providing artifacts* were reconceptualized as existing in the center of the model but then extending throughout the model to be associated with *extending evaluation use*. And, as previously mentioned, the theme of *additional audiences* emphasized in Phase I was subsumed into this theme as merely a component of *extending evaluation use*.

In keeping with the simplification required to illustrate a conceptual model, Figure 16 visually associates (3) *interact (connect)* with (5) *instrumental use* and portray (4) *remember (learn)* as most directly related to *conceptual use* and (6) *extending evaluation use*. The link between *remember (learn)* and *conceptual use*, in which knowledge and understanding has transferred into the audience and is available for future decision-making is obvious, but participants described (6) *extending evaluation use* as far beyond just *conceptual use*. In this broad conceptualization of use, participants were considering future potential relationships and the entire field of evaluation.

It is not a surprise that professional evaluators emphasized long-term – sometimes very long-term – relationships, or that they extended the satisfaction associated with audience engagement into the positive reputation of their firm. As service-based professionals relying on future contracts, it follows that each relationship may lead to other relationships in the distant future, such that a positive experience “maybe makes you think better of [our company] or the work that we're doing or whatever we're engaged in, and that might not come to fruition for years or decades” (VV-17).

What was a surprise to me was that four of the six specifically discussed their role in terms of the entire field of evaluation. It was clear that these experienced evaluators were thinking beyond their own financial and reputational needs to consider evaluation as a field with comments such as “if [the audience] uses the skills they have built through the evaluation process, that's a win for me, professionally, regardless of whether or not you ever say my name ever again. It's a win, for the field, right?” [VV-17]. Promoting the field of evaluation appeared to be intentional, prompting me to ask one interviewee, *Do you consider yourself a cheerleader for evaluation as a field?* When framed this directly, the answer was yes. Unfortunately, this conceptualization of *extending evaluation use* was not fully explored due to the time and scope limitations of this project.

Phase II: Limitations

Relying on follow-up interviews using the general contours of grounded theory was appropriate for the limited goal of this project, but the number of interviews conducted for this project was insufficient for a true grounded theory approach. This phase did yield the additional information necessary to sufficiently describe the model of *audience engagement* for inclusion of Phase III, but some aspects of the model remain ill-defined. For instance, later interviews began describing the evaluator specific activities of the model as inputs and outputs. This framing would likely resonate with evaluators and could have been pursued through additional interviews. It was simply unknown at the beginning of this project that such a rich model would emerge and properly describing it is outside the scope of this project.

Grounded theory relies on the flexibility of theoretical sampling (Corbin & Strauss, 2015), whereby the researcher can target the participants who are best able to address the theory gaps. The time limits associated with a doctoral project as well as the IRB limitation of follow-up interviews meant that the flexibility of theoretical sampling was limited to the eighteen participants from Phase I. Still, careful memoing and organization of these memos allowed for strong analysis even considering the sample limitation.

"Data saturation" is associated with rigor in qualitative research, but the ability to achieve saturation depends on both the definition of saturation being used and the homogeneity of the interview sample (Sim et al., 2017). In this case, my definition of saturation is based on meaning rather than presence of themes or codes, but I do not know how homogenous my sample was. Participants were identified as data viz experts and evaluators, but the grounded theory approach in Phase II used the same sample as Phase I out of time and IRB expediency consistent with a doctoral project. A true grounded theory approach would have allowed interviews with data viz experts in evaluation outside the original sample. Phase III will provide some insight into the model from other evaluators, but the fact remains that this is not a fully formed model and further research effort is needed to fully describe it. If Phase III suggests that this model is useful to evaluators, then it will warrant future effort.

One potential weakness is that I recognized the overlap between the first four buckets of the emerging *audience engagement* model and the Information Processing Model of Memory as described by Huang and colleagues (2009) very early in my analysis. After that association had been made, all subsequent comments appeared to back up or reinforce that model, thereby confirming my bias that the audience engagement model relies on the Information Processing Model. Typically, a grounded theory approach seeks to allow a theory to emerge on its own from the direct words of the participants and only later – after the analysis has been completed – to attempt to locate that new theory "within the larger body of professional theoretical knowledge" (Corbin & Strauss, 2015). While a concern worth noting, this is not significant enough to stop developing the model. That the *audience engagement* model has – at its core – strong overlap with a well-known model applicable to nearly any field lends credibility to the model. It also allows the reader to de-prioritize the obvious aspects of the model and consider the far more interesting evaluation specific portions of the model.

Phase II: Conclusion

A model of audience engagement was identified in Phase I and then described in Phase II well enough to (1) visually depict the model and (2) describe it in a four-minute explainer video to be used in Phase III. However, it was simply unknown at the beginning of this project that such a rich model would emerge and properly describing it is outside the scope of this project. While further research is needed, there is still an opportunity to learn more about the emerging model in Phase III when it will be shared with a sample of practicing evaluators.

There was a great deal of model development and reconceptualization between Phase I and Phase II presentations of the model as readily apparent when comparing Figure 11 and Figure 16. Some of the most important and evaluation specific portions of the Phase II model were not identified during

Phase I analysis. The amount of development that occurred through just six interviews suggests that an even more robust grounded theory approach was warranted and suggests that the Phase II model presented is far from "complete." This is being presented as a model in Phase III, but it most correctly should be considered an emerging or nascent model needing further research and development.

The *audience engagement* model is a complex model that touches on multiple aspects of evaluation, regardless of the inclusion of data viz. As with so much else in this model, data viz is merely one component of the *audience engagement* being discussed. I am presenting data viz and the information processing model of memory as a core around which to wrap this model of *audience engagement*, but it is possible that a model of *audience engagement* exists separately from this data viz discussion. While it is clear that data viz can play a role in encouraging and extending evaluation use, I am not claiming that data viz accomplishes the desired outcomes of *evaluation use* on its own.

Phase II: Future Directions

Now that the *audience engagement* model has been sufficiently described, it will be articulated into a brief explainer video for use in Phase III. The *efficiency rationale* of data viz, the *explain* ←→ *explore* model, the *data* → *insight* model, and the *audience engagement* model will be presented to a sample of evaluators in Phase III. With the goal of better understanding the potential role for these conceptual models in evaluation, evaluators will be asked if they are familiar with the models and the extent to which they find them appropriate and useful in evaluation. The models will also be analyzed for their strengths, weaknesses, and usefulness to better understand the areas of overlap and suggest potential integrated model(s).

Phase III

Phase I relied on a secondary analysis of existing data to determine whether the underlying *efficiency* argument for data viz *use* as well as the three proposed models of data viz *use* were present within interviews of eighteen practicing evaluators who are experts in data viz. Findings suggest that the *efficiency* rationale and all three models (*U-FE Framework, explore* ←→ *explain*, and *data* → *insight*) were present to varying extents, see Phase I and Figure 12 for details.

Phase II used a grounded theory approach, (Corbin & Strauss, 2015; Taylor et al., 2016) with follow-up interviews to better understand and articulate how practicing evaluators conceptualize an *audience engagement* model of data viz.

Phase III will present the *efficiency rationale* of data viz and three conceptual models (*explain* ←→ *explore*, *data* → *insight*, and *audience engagement*) to a sample of evaluators to determine whether these models are considered appropriate and useful in evaluation.

Phase III Research Question

Phase III is framed as seven research questions, divided into three areas, seeking to understand if evaluators accept the underlying *efficiency* rationale, if there are significant differences between the models, and the perceived strengths, weaknesses, and uses of each model.

Efficiency Rationale of Data Viz

1) (QUANT) To what extent are participating evaluators (a) familiar with and (b) accepting of the increased *efficiency* rationale?
2) (QUANT) To what extent do participating evaluators consider each model "adds value" to the *efficiency* rationale?

Directly Comparing the Models

3) (QUANT) Are there significant differences between the models in evaluator (a) familiarity, (b) acceptance, (c) perceived appropriateness in evaluation, and/or (d) perceived usefulness in evaluation?
4) (QUANT) Did evaluators "learn something new" from watching the model videos?

5) (QUANT) Are there significant differences between the model videos as demonstrated by (a) survey exit rates following a video or (b) evaluator model preferences after watching multiple videos?

General Analysis of the Models

6) (QUAL) Are model improvements suggested by this sample of evaluators?
7) (QUANT/qual) For each model under consideration, (a) to what extent and (b) in what ways, have participating evaluators *used* the model, and (c) what outcomes do they attribute to that *use*?

Phase III: Method

Phase III is the final phase of a larger exploratory multi-phase mixed-methods study (Creswell & Poth, 2018). This final phase relies primarily on quantitative methods to compare awareness and perceived usefulness of these models of data viz *use* through a Qualtrics survey of evaluators. Open-ended qualitative feedback is used to better understand the quantitative data collected and suggest model improvements for future work. Claremont Graduate University certified this project exempt under IRB # 4655.

Obtaining the Sample

As an unexpected and significant issue arose when distributing the survey to the intended sample, the general survey creation and distribution method will be discussed here to better explain how the sample was achieved. Description of participants and materials, data handling and cleaning, data analysis, etc. will be discussed under traditional headings later.

Piloting the Survey

A draft survey was shared with research lab members (see Appendix A) in early fall 2023. Survey flow improvements were made based on their input (e.g., language on several questions was simplified, the request for demographic information was moved to the end of the survey, etc.). I then piloted the updated survey to ten CGU evaluation students and faculty in late fall 2023. Most comments at this point were suggesting edits to the videos themselves, indicating that the survey was ready for distribution.

Distributing the Survey

After formal review of my request, AEA provided a list of 2,000 current member emails. I removed one duplicate email, one test account, and the three contacts that had participated in interviews. I sent the approved email wording (Appendix H) to 1,995 individuals via Qualtrics using the internal Qualtrics email system. After six days, I had obtained an abysmal response rate of only 17 (0.85%) completed surveys. Investigation showed that only 540 (27.07%) emails had been opened after 6 days. Critically, zero of the opened emails were Gmail accounts even though 539 (27.02%) of the total list were Gmail accounts. Comparing this 0.00% open rate for Gmail accounts to the 37.09% open rate of all other email accounts strongly suggests that Qualtrics emails to Gmail accounts were blocked by Gmail servers prior to delivery. (Note that it is unknown how many of the "all other" email accounts were administered by Gmail and also blocked. A colleague on the distribution list using a custom email administered by Gmail confirmed that the survey email generated by Qualtrics was not delivered to her. It is, therefore, likely that a number of other unopened emails are from accounts administered by Gmail.) In addition to effectively decreasing the list size and understating the response rate, another negative consequence of failing to deliver requests to smaller eval firms and sole proprietors is that the data would be biased toward university and large firms hosting email on their own servers.

To combat this for the allowed reminders, I exported the distribution list from Qualtrics including the unique survey link for each participant. It was important to use their individual link to allow Qualtrics to update its internal metrics regarding open and completion rates as well as to ensure the user experience such that those already in progress with the survey would be returned to the point they left off and those who had completed the survey would not be able to submit again. I removed the 17 participants who had completed the survey and the 31 emails that had been bounced by their respective servers. I then imported that list of 1,948 emails into GMass and sent individualized emails – including their unique survey link – to those individuals. GMass is a fee based bulk email service integrated into Gmail that automates sending individual emails to large lists. The campaign took five days to avoid Gmail sending limits and spam filters. However, the large increase in email volume still triggered Gmail to close my account and it was only reinstated after a formal request. The list was cleaned again using similar procedures the following week and one more round of reminders were sent via GMass. This resulted in fewer reminders to Gmail accounts, which received two requests to complete the survey, while all others received three requests, as permitted by AEA list rules.

Survey Response Rate

Considering the 5 emails I removed and the 43 undeliverable emails, the actual list size is 1,952. I did remove additional emails from reminder rounds (e.g., people who emailed me back saying that they couldn't complete it, auto-replies indicating that the person could not complete the survey, etc.). These removals were to avoid being identified as a spammer, not for research purposes, so those removals are not reflected here.

The 134 surveys submitted from 1,952 emails is a response rate of 6.9%. Upon reviewing the data, two more surveys were removed from this analysis: One participant declined to participate and one participant responded "not an evaluator" even though they were clearly a member of AEA and comments suggest the participant is an analyst at an eval firm. This left 132 surveys (6.8%) for analysis. One additional participant was removed during the data cleaning process, as will be described later.

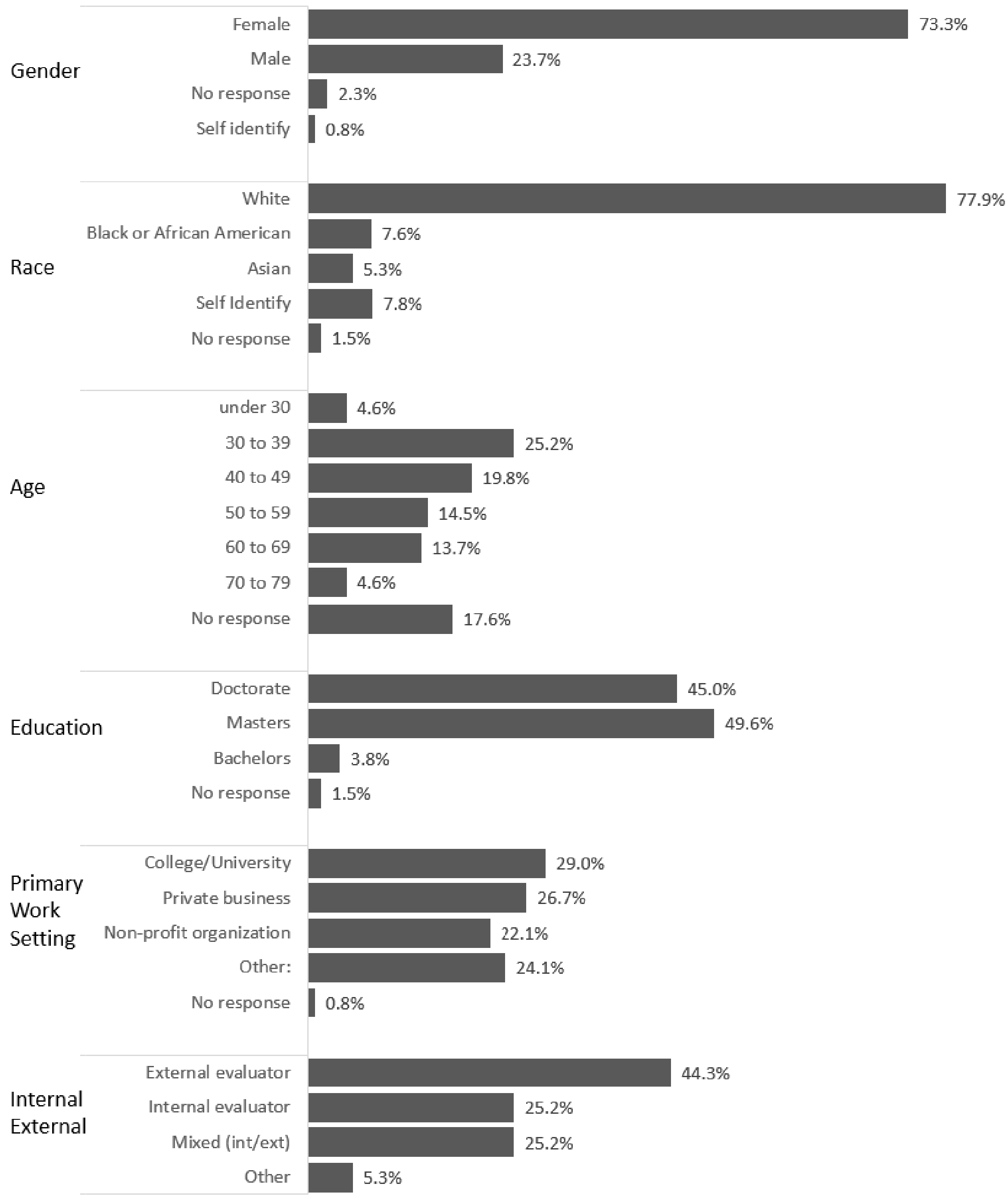
Phase III Participants

Participants in this phase were 131 evaluators recruited from the membership of the American Evaluation Association (AEA). Unique survey link codes ensured that participants could only participate once. Absolutely no deception was used in this study. Participants were informed via email (Appendix H) that they were invited to view brief, evidence-based, explainer videos on different ways of considering data viz *use* in evaluation. They were provided with informed consent (Appendix I) at the beginning of the survey (Appendix J) and permitted to exit the survey at any time.

Figure 19 shows that participants were primarily female (n = 96; 73.3%), White (n = 101; 77.1%), and held graduate degrees (n = 124; 94.7%). The majority of participants (n = 74; 56.5%) identified as White and female, with a graduate degree. Participant age (M = 46.8, SD = 13.7) received the lowest response rate of all demographic items, with 16.8% (n = 22) choosing not to respond. The majority of participants were fulltime evaluators (n = 101; 76.5%) and native or fluent English speakers (n = 118; 90.1%).

Figure 19

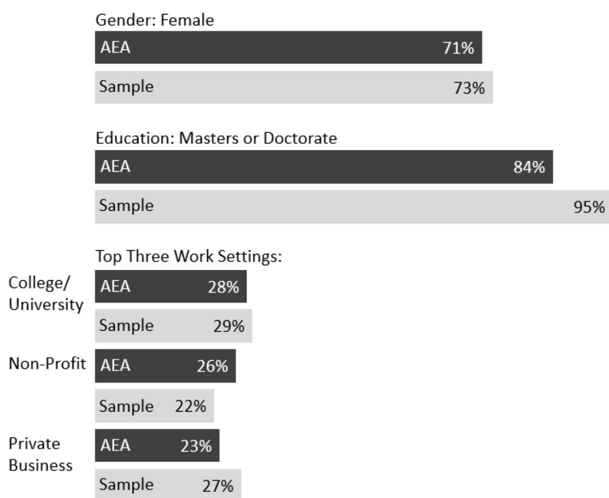
Phase III Participant Demographics



Through personal communication, AEA provided demographic comparison data for membership year 2023. Figure 20 shows that the sample of evaluators collected for this study is similar to the AEA membership population on the demographic characteristics provided. The only notable difference, that the sample had a higher percentage of participants with a master's degree or higher can be explained by the 11% of AEA participants not disclosing their education compared to only 1.5% of sample participants not disclosing their education level.

Figure 20

Comparison of Sample Demographics to AEA Membership Demographics



Phase III Materials

The materials used in Phase III were four explainer videos and a survey delivered via Qualtrics.

Explainer Videos

While conducting Phase II analysis, I created short explainer videos describing the *efficiency rationale* of data viz, each of the three previously described conceptual models (*U-FE Framework, explain ←→ explore*, and *data → insight*), and the *audience engagement* model. These videos were reviewed by data viz experts and some changes were made, including removing one of the models from this research project.

Videos Were Created. A script for each video was created based on descriptions of each model from the literature review and from this research project. Videos were created in Microsoft PowerPoint with heavy reliance on animation features in that software. Unfortunately, the audio recordings in

PowerPoint are saved per slide, so audio was lost during slide transitions. Therefore, the video content was created in PowerPoint and then Loom.com screen recording software was used to capture and edit the screen and audio recording. These edited recordings were then uploaded to YouTube and their unlisted links were shared with the data viz expert reviewers.

Videos Were Reviewed and Edited. In addition to review by my faculty advisor, videos were reviewed and edited by three members of the Data Visualization Research Lab at Claremont Graduate University. All members of the lab are advanced graduate students in evaluation with expertise in data visualization. This team is also familiar with the data, having conducted and analyzed the 2021 interviews used in Phase I (see Appendix A). Editing suggestions were made via email and during weekly meetings. Edits were made, videos were re-recorded, and additional review was held in an iterative manner. Most editing suggestions improved wording, examples, and visual depictions, but two decisions that arose from this review process had a significant impact on this research project and are described below.

The U-FE Model Was Removed from This Research Project. As described in the literature review, I created this model by applying the existing *Utilization-Focused Evaluation* Model (Patton, 2008; Patton & Campbell-Patton, 2022) to data viz. The *U-FE* model itself is an involved model about which entire books have been written. A great deal of time is required to explain the differences between *findings use* and *process use* and then explain each of the use types (*instrumental*, *conceptual*, and *symbolic*) within those two categories. Then each of those *use* types must be explained within the context of data viz with examples. The reviewers suggested that the draft video was too long and conceptually dense for inclusion in this project. Also, from a purely practical perspective, reducing the number of videos reduced the responses required in the Phase III survey as the analytical method planned (i.e., directly comparing the models with Chi-square) requires at least ten Phase III participants per video included.

Phase I supported that – at least that sample of – practicing evaluators do extend the *U-FE* conceptualization of *use* to data viz in their work. However, the plan for this project was to present the relevant models as brief explainer videos and the *U-FE* conceptual model applied to data viz did not meet the brief requirement. Therefore, the *U-FE* model was removed from this project; it is not presented in Phase III based on practical limitations and not due to any weaknesses in the model itself. It likely is an important conceptual model of data viz in evaluation and future research could pursue a better description of and greater understanding of this model.

The Audience Engagement Model Was Simplified. The *audience engagement* model that emerged in this research clearly references the Information Processing Model of Working Memory (Huang et al., 2009) common in social sciences literature. The importance of this foundational model is described in this research project and depicted in the model as presented in Figure 16. However, the visual depiction (e.g., the gray boxes) and direct references to “information processing” and “working memory” were removed from the video visual and script. The important aspects of the models (e.g., *attract attention*, *hold attention*, *interact* (*connect*), and *remember* (*learn*)) are still presented and explained in the video. This change reduced the video run time by nearly 40 seconds and allowed greater emphasis on the evaluation specific aspects of the model, which I considered more integral to this research.

Final Videos Used in Phase III. The three model videos ranged from 184 to 237 seconds in length (M = 218; SD = 24.1) The *Introduction to Data Visualization Efficiency* video was 182 seconds long and viewed by all participants. Participants were then provided one of the three model videos at random, see Figure 21 for survey flow, and asked questions about that model (see Appendix I for questions). Participants were then presented with the option to end the survey and complete their demographic data or to continue and watch another video. Participants choosing to continue were shown another model video, asked the same questions about the new model, and then given another opportunity to exit or continue. Final video titles, run times, and links to watch them are:

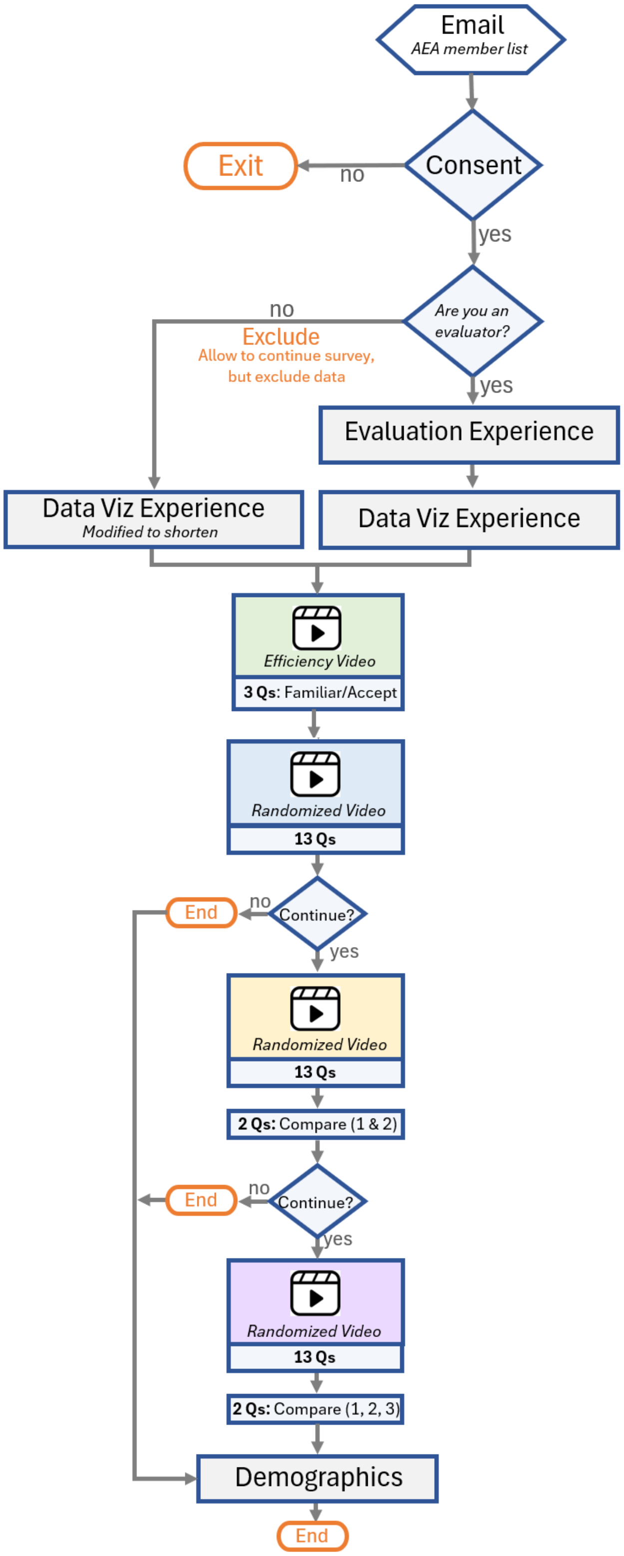
- *Introduction to Data Visualization Efficiency* (3:02) (https://youtu.be/yUiDYoyaIlU)
- *Explain or Explore Model* (3:04) (https://youtu.be/GN2BILFwagQ)
- *Data to Insight Model* (3:57) (https://youtu.be/teaZjy3Ybr8)
- *Audience Engagement Model* (3:53) (https://youtu.be/agyBy3Ek3Jg)

Survey

After obtaining consent (Appendix I), the first survey question asked if the participant is an evaluator. To encourage honesty, curious non-evaluators were permitted to watch the videos and participate in data collection but they were informed that their responses would not be used in this research. Evaluator participants were asked to rate their *evaluation knowledge and skill*; their *data viz knowledge and skill*; their ability to *understand statistics*, *use charts and graphs to make decisions*, *create data viz*; and the frequency with which they *see*, *use*, *create*, and *learn about* data viz. Exact question wording and response sets are available in Appendix I. Evaluator participant demographic data was collected consistent with the AEA demographic questionnaire to allow comparison of this sample to the AEA membership. Data related to evaluator experience and data viz experience were collected before the videos were presented to minimize time required to complete the survey after they declined to watch more videos. Demographic data such as race, gender, and age were collected at the end of the survey. Figure 21 offers a visual flow of the survey process.

Figure 21

Survey Data Collection Flow



All participants viewed the video explaining the increased *efficiency* rationale of data viz *use* and were asked to indicate their level of agreement with the following statements on a 1 to 5 Likert-type scale of *Strongly Disagree* to *Strongly Agree*:

1) Data visualization can be used to increase efficiency over text-based communication.
2) Before I watched the video, I was familiar with the efficiency rationale for data viz.
3) The efficiency rationale for data viz makes sense to me.

The goal was to ascertain whether participants accept the underlying rationale of increased efficiency and gauge their familiarity with it.

Participants were then provided a brief explainer video (less than four minutes) on one of the conceptual models (*explain* ←→ *explore*, *data* → *insight*, or *audience engagement*) at random. After viewing the video, participants were asked to indicate their level of agreement with the following statements on a 1 to 5 Likert-type scale of *Strongly Disagree* to *Strongly Agree* to discern how familiar and useful they find the conceptual model:

1) Before I watched the video, I was familiar with the [model] way of thinking about data viz.
2) I have learned something new about the [model] model from watching this video.
3) The [model] model relies on the concept of increased efficiency.
4) If used, the [model] model adds value to the "efficiency rationale" for using data viz.
5) The [model] way of thinking about data viz makes sense to me.
6) The [model] way of thinking about data viz aligns with my personal approach to evaluation.
7) The [model] way of thinking about data viz is appropriate in evaluation.
8) The [model] way of thinking about data viz would benefit stakeholders.
9) I have used the [model] model to *describe* my reason for using data viz to others.
10) I am likely to use the [model] way of thinking about data viz in future evaluation work.

Participants were also asked with what frequency they have used the model (i.e., *I have used the [model] model in my evaluation work*) on a scale of *Never* (1), *A few times* (2), *Often* (3), *Almost Always* (4), or *Always* (5). If they had used the model, they were asked: *please describe one or two outcomes from using the [model] in your evaluation work [text entry].* Finally, the participants were asked to provide text entry regarding what they perceive as the *strengths* and *weaknesses* of the model.

Upon completion of the survey questions, participants were thanked for their time and told they were done with the study but were welcome to watch another explainer video and respond to the same questions. Participants exiting at this point were taken to the demographic questions. Participants choosing to continue were shown another model video, asked the exact same questions, and then given another opportunity to exit or watch another video.

Participants electing to watch a second and/or third video were presented with a list of the models they had viewed and asked two additional questions:

Please compare this model to the other model(s) you have learned about during this survey.

1) Which model is most useful to your evaluation work? [select from models presented]
2) Why did you answer the way you did? [text entry]

After watching one, two, or three model videos, participants were asked to provide demographic data (no questions were required) and then thanked for their time.

Phase III Data Analysis

Consistent with a project on data visualization, data visualization was used extensively in the preliminary analysis to search for patterns and relationships in the data. Microsoft Excel and Microsoft PowerPoint were primarily used for this analysis. Traditional statistical analyses were conducted in SPSS version 29.0.2.0.

Quantitative Data Cleaning, Initial Data Inspection, and Removal of Outlier

All data were exported from Qualtrics into Microsoft Excel. Data cleaning and organization was completed in Microsoft Excel. Missing demographic data was re-coded "Choose not to respond" as all questions had been presented. Some participants were removed case-wise, as discussed, but no missing data were imputed. All missing data will be identified as such in reporting.

Participant age was calculated from the data field "year of birth." Numeric data that had been collected open-ended were converted to new data fields and calculated as conservatively as possible. For instance, "number of completed evaluations" was an open text field with responses ranging from "zero" to "hundreds;" conservative recoding resulted in "hundreds" being calculated as 200, "more than a hundred" as 100, "more than 10" as 10, etc. An additional variable, "First," was created for each model question so that data from each participant's "first viewed model" could be analyzed independently and allow direct comparison of the models.

Initial review of all data for 132 participants was done via review of descriptive statistics and analysis of Mahalanobis distance, and then through visual inspection of histograms and the analysis of outliers feature of SPSS.

Considering that most data in this phase were collected on five-point Likert-type scales from *Strongly Disagree* (1) to *Strongly Agree* (5) with *Neutral* (3) or as ordered data (e.g., 1 = *Never*; 2 = *Less than once a month*; 3 = *Once a Month*, 4 = *Once a Week*; and 5 = *Once a Day or More*), the original analysis plan intended to consider the data as ordinal and rely on the limited number of statistical tests available for such data. However, although Shapiro-Wilkes test of normality was significant at $p < .001$ for all variables (see Appendix K), only 9 out of 26 (34.6%) skew and 5 out of 26 (19.2%) kurtosis scores were outside the pre-established cut-off of 1 and 3, respectively. Calculation and review of p-value of Mahalanobis distance suggested that 11 subjects had extreme outlier data points.

Using the analysis of outliers feature of SPSS along with visual review of histograms and case by case analysis of responses, each variable was reviewed again with special attention to those violating skew or kurtosis cut-offs, as well as the specific 11 subjects flagged for review. Upon review, most of the violations had a logical explanation. For instance, SPSS suggested that there were 3 outliers who had completed greater than 150 evaluations. All three participants were among those flagged for review, but their responses and other data were kept under the assumption that the numbers provided were their honest responses. The only data points that did not fit a logical pattern were from one participant who responded *strongly disagree* on two questions to which no other participants responded *strongly disagree*, resulting in the most egregious kurtosis violations of the data (16.67 and 7.62). The participant also left more questions blank than any other respondent and supplied no qualitative comments on the one model viewed. This participant was removed from analysis. With all remaining violations being less than extreme and having plausible explanations, data analysis continued with 131 participants.

Treating Ordinal Data as Continuous

Only three quantitative data points in this study (age, years of evaluation experience, and number of completed evaluations) can be considered truly continuous variables. Most demographic data points (e.g., race, gender, etc.) are categorical, as well as the variable "first model video watched." (e.g., EE = *explain* ←→*explore*, DI = *data* → *insight*, and AE = *audience engagement*). The original analysis plan was to consider the remaining data, which were collected on five-point Lickert-type scales, as ordinal data. However, as discussed above, after removing just one outlier, skew violations were reduced to 6 (with only 1 variable above 2) and there were only 3 kurtosis violations of 3.05, 3.46, and 7.69. The one variable with known extreme outliers, "number of completed evaluations," accounted for both the highest skew of 2.45 and the kurtosis of 7.69, see Appendix L.

After careful inspection, consultation with a statistician, and thorough consideration, these variables will be treated as continuous in this analysis. This is consistent with other research presentations of similar data and relies on the common knowledge that ANOVA is robust to minor deviations from normality, especially with larger sample sizes. Assumptions for each of the ANOVA and chi-square tests will be detailed.

Nonetheless, I am mindful of the assumptions underlying this analysis approach and do not want to argue that it is philosophically acceptable to treat ordinal data as continuous, particularly in evaluation. I will, therefore, present the data visually in a manner consistent with both a continuous assumption (i.e., with mean and standard deviations) as well as an ordinal approach (i.e., grouped by response).

Relationships Between Variables

Figure 22 shows the Pearson correlation coefficient between continuous variables; missing data were removed pairwise. Using the general guidelines provided by Cohen (1988), strengths of association are presented as light green for a small correlation ($0.1 < | r | < .3$), medium green for a moderate correlation ($0.3 < | r | < .5$), and dark green for a strong correlation ($| r | > .5$).

Figure 22

Pearson Correlation Matrix of Variables

	Experience			Knowledge/Skill		How Easy or Hard			Monthly Frequency			
	years	# evals	Exp viz freq use	eval	data viz	unstd stats	unstd graphs	create graphs	see charts	use charts	create charts	learn viz
Exp years	--											
Exp # evals	0.349***	--										
Exp viz freq use	-0.134	0.110	--									
Rate eval KS	0.484***	0.325***	0.013	--								
Rate viz KS	-0.071	0.060	0.521***	0.132	--							
unstd stats	0.130	0.174	0.186*	0.246**	0.095	--						
unstd graphs	0.198*	0.291**	0.238**	0.178*	0.226**	0.529***	--					
create graphs	0.087	0.221*	0.278**	0.204*	0.343***	0.359***	0.534***	--				
see charts	-0.151	0.061	0.361***	0.041	0.292**	0.341***	0.201*	0.248**	--			
use charts	-0.117	0.028	0.234**	0.118	0.341***	0.303***	0.125	0.251**	0.512***	--		
create charts	-0.156	0.174	0.356***	0.025	0.376***	0.195*	0.183*	0.331***	0.547***	0.581***	--	
learn viz	-0.247**	0.033	0.148	-0.094	0.242**	0.040	-0.028	0.131	0.266**	0.413***	0.578***	--
increased efficiency	-0.066	0.143	0.312***	-0.079	0.194*	0.000	0.147	0.107	0.127	0.001	0.152	0.047
familiar before	0.111	0.078	0.100	0.130	0.306***	0.082	0.195*	0.172*	-0.024	0.117	0.144	0.085
makes sense	-0.013	0.193*	0.161	0.009	0.023	-0.065	0.173*	0.029	-0.044	-0.145	-0.011	-0.083
familiar before	0.097	0.092	0.161	0.184*	0.328***	0.078	0.138	0.203*	0.146	0.234**	0.215*	0.077
learned new	-0.139	0.028	0.068	-0.057	-0.031	-0.043	-0.114	-0.085	0.118	-0.039	0.021	-0.044
described others	-0.006	0.019	0.192*	0.074	0.248**	0.019	-0.063	0.147	0.068	0.243**	0.267**	0.187*
likely to use	-0.169	-0.039	0.092	-0.172*	-0.046	-0.045	-0.048	0.051	0.078	0.058	0.120	0.107
makes sense	0.045	0.197*	0.197*	0.087	0.215*	0.088	0.173*	0.178*	0.158	0.140	0.179*	0.091
appropriate	-0.026	0.052	0.155	-0.089	0.137	-0.094	0.037	0.007	0.016	0.065	0.043	-0.054
aligns	-0.096	0.052	0.280**	-0.140	0.126	-0.004	0.058	0.111	0.074	0.105	0.118	0.102
benefit	-0.121	-0.053	0.189*	-0.111	0.101	-0.037	0.047	0.076	0.176*	0.143	0.076	0.104
adds value	-0.130	0.003	0.156	-0.047	0.187*	-0.067	0.109	0.106	0.072	0.087	0.081	0.057
have used	0.100	0.096	0.369***	0.168	0.334***	0.063	0.091	0.160	0.130	0.185*	0.217*	0.068

Pearson Correlation Matrix Continued

	3Qs Efficiency			First Model Video									
	increase efficiency	familiar before	makes sense	familiar before	learned new	described others	likely to use	makes sense	approprt	aligns	benefit	adds value	have used
increased efficiency	--												
familiar before	0.120	--											
makes sense	0.665***	0.179*	--										
familiar before	0.172*	0.368***	0.055	--									
learned new	0.288**	-0.154	0.219*	-0.124	--								
described others	0.043	0.202*	-0.030	0.645***	-0.059	--							
likely to use	0.112	0.026	0.057	0.231**	0.385***	0.429***	--						
makes sense	0.358***	0.129	0.321***	0.243**	0.254**	0.215*	0.322***	--					
appropriate	0.323***	0.040	0.270**	0.166	0.388***	0.161	0.454***	0.764***	--				
aligns	0.232**	0.127	0.187*	0.142	0.279**	0.228**	0.592***	0.560***	0.710***	--			
benefit	0.217*	-0.038	0.167	0.219*	0.328***	0.250**	0.538***	0.521***	0.635***	0.667***	--		
adds value	0.222*	0.001	0.267**	0.288**	0.324***	0.276**	0.406***	0.517***	0.537***	0.418***	0.710***	--	
have used	0.077	0.156	0.036	0.570***	-0.061	0.640***	0.368***	0.235**	0.241**	0.273**	0.249**	0.232**	--

Note. * *p* < .05, ** *p* < .01, *** *p* < .001. Significant negative correlation is presented in red font.
Strength of significant relationships is color shaded green with darker green indicating a higher correlation.

Directly Comparing the Models

As each participant can respond to multiple models, only the first response set on each can be directly compared between subjects. This is to ensure the independence of the samples, an important assumption in both ANOVA and chi-square tests. Two other foundational assumptions of the study, random assignment and balanced design, are met, as detailed in Appendix M. Randomization of video watch order was confirmed via non-significant chi-square tests of independence for all categorical variables and a chi-square goodness of fit calculation confirms that the number of participants randomly assigned to each of the three conditions are equal as the difference between each and the expected value of 43.7 is not significantly different ($X^2(2) = 0.20$, $p = .906$). Appendix N details the assumptions required for the chi-square goodness-of-fit test and chi-square test for association used.

Levene's test for equality of variances for two questions: familiar before watching ($p < .001$) and makes sense ($p = .003$) determined that the assumption of homogeneity of variances was violated. Therefore, a Welch’s ANOVA with Games-Howell post hoc was used to directly compare the three models on the ten questions of interest, as recommended by Laerd Statistics (2024). Assumptions made for this analysis are detailed in Appendix O.

Conventional Content Analysis for Qualitative Analysis

Qualitatively, each model was assessed to better understand its perceived *strengths* and *weaknesses* within evaluation, opportunities to *improve* the model and/or combine it with other model(s), and its perceived *usefulness* in evaluation.

Each open-ended comment question was asked in a way intended to directly answer the associated research question (e.g., participants were asked “I think the *strengths* of the [model] are:” separately from “I think the *weaknesses* of the [model] are:”). With no *a priori* codes for any of these topics, I used a conventional content analysis approach where coding categories were derived directly from the data (Hsieh & Shannon, 2005).

Data cleaning and word count calculations were performed in Microsoft Excel. I removed responses such as “none,” “everything you said!” and even “I really don't think there are any weaknesses with the model” that didn’t add value to analysis. I also removed responses that referred to previous responses such as “see above” since those comments will be themed and discussed with their associated question. It is the nature of this data, and a significant challenge in this analysis, that there is tremendous overlap between the models.

Data analysis on model *strengths*, *weaknesses*, and *usefulness* was conducted by thoroughly reading all comments on one question for one specific model several times and then beginning to apply codes. Codes were merged and refined through multiple readings. As I was focused on intended meaning within the context of the specific model, codes are not necessarily comparable across models. For instance, one *strength* code from the *explain* ←→ *explore* model, "simplicity of the model (n = 7)" emphasizes that the actual simplicity of the model – the fact that it only has two categories – makes it easy to understand. There is a *strength* code for the *data* → *insight* model with overlapping meaning as participants indicated that the model is “intuitive (easy to understand) for evaluators and/or stakeholders (n = 11)” but this code associates “easy to understand” with the cognitive science underpinnings of the model and is not strictly equivalent to “simple.”

The exception to this approach was on the question that asked participants to directly compare the models after watching two or three videos. Here, their comments are considered in the context of two or three models at once.

Phase III: Ethical Considerations

No deception was used in this study. One intent of this study was to provide knowledge in exchange for participant time. Out of respect for participant time, I endeavored to keep the time required to complete the survey questions as short as possible.

Phase III: Findings

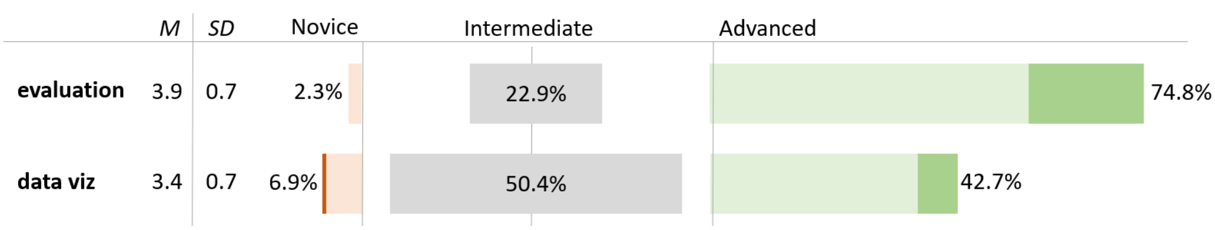
After presenting findings related to general participant experience in evaluation and data viz, the research questions will be addressed in order.

Participant Experience in Evaluation and Data Viz

Participants reported having an average of 14.8 years of evaluation experience (*SD* = 10.1; *Range* = 0 – 50] and 45.6 completed evaluations (*SD* = 52.8; *Range* = 0 – 300]. Figure 23 shows most participants reported they were advanced (55.0%) or expert (19.8%) in evaluation, but less than half reported that they were advanced (35.9%) or expert (6.9%) in data viz. There is no correlation between these responses (*r* = 0.132, *p* = .132).

Figure 23

Participant Knowledge and Skill in Evaluation and Data Viz (N = 131)

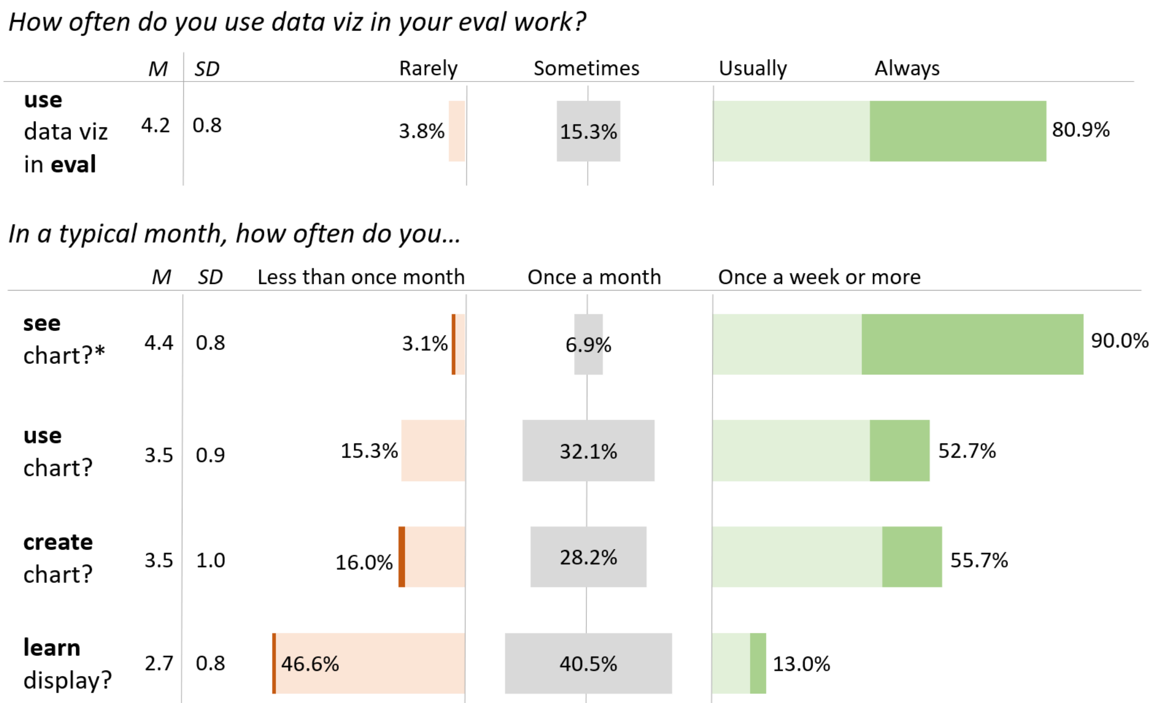


Note. Responses are grouped as low (*none* and *novice*) to the left in shades of orange, intermediate responses in gray down the middle, and high responses (*advanced* and *expert*) to the right in green. Lowest (*none*) and highest (*expert*) responses are presented with more saturated colors in the bar. The numeric percent displayed outside the bar is the combination of low or high responses.

Figure 24 shows that most participants *usually* (38.2%) or *always* (42.7%) use data viz in their evaluation work (*M* = 4.2; *SD* = 0.8) and provides greater detail into the ways in which they use data viz.

Figure 24

Participant Frequency of Using Data Viz in Their Work (N = 131)



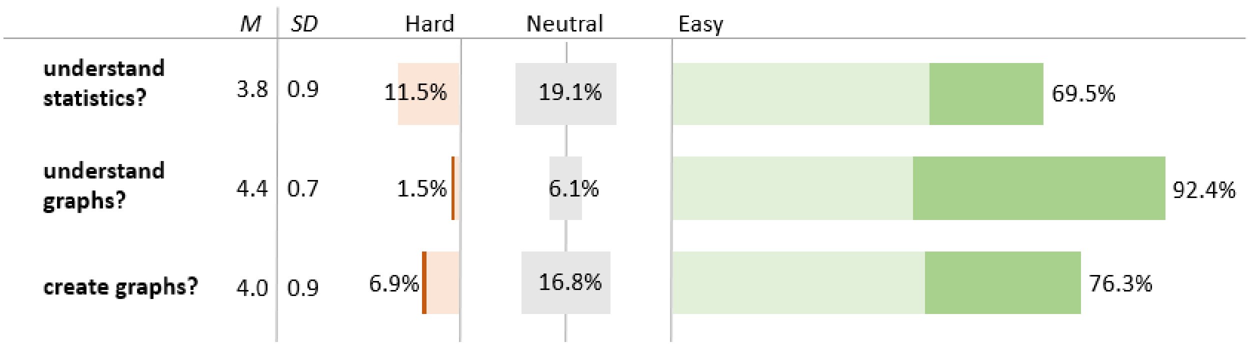
Note. Participants were asked in general how often they use data viz in their eval work with options from (1) Never to (5) Always and then how often they **see** data in a chart, **use** data in a chart to make decisions, **create** a chart, and **learn** new ways of displaying data. Options were (1) Never, (2) Less than once a month, (3) Once a month, (4) Once a week or more, or (5) Once a day or more. Lowest (*Never*) and highest (*Once a day or more*) responses are presented with more saturated colors in the bar. *This question had one missing response (*N* = 130).

Figure 24 shows that participants frequently see data presented in charts (*M* = 4.4; *SD* = 0.8) but are less likely to use data viz to make decisions (*M* = 3.5; *SD* = 0.9) or create a chart (*M* = 3.5; *SD* = 1.0). Participants *use* data viz and *create* data viz at approximately the same rate. Using and creating data viz are also strongly positively correlated ($r = 0.58$, $p < .001$), suggesting that those who use are also creating data viz. Learning new ways of displaying data was the least common activity of the four presented with the majority of participants (87.1%) doing so once a month or less.

Figure 25 shows that participants generally found it easy to understand (*M* = 4.4; *SD* = 0.7) graphs and charts, but slightly less easy to create them (*M* = 4.0; *SD* = 0.9). Their lowest ease was with understanding statistics (*M* = 3.8; *SD* = 0.9).

Figure 25

Participant Comfort with Data Viz: "In general, how easy or hard do you find it to..." (N = 131)



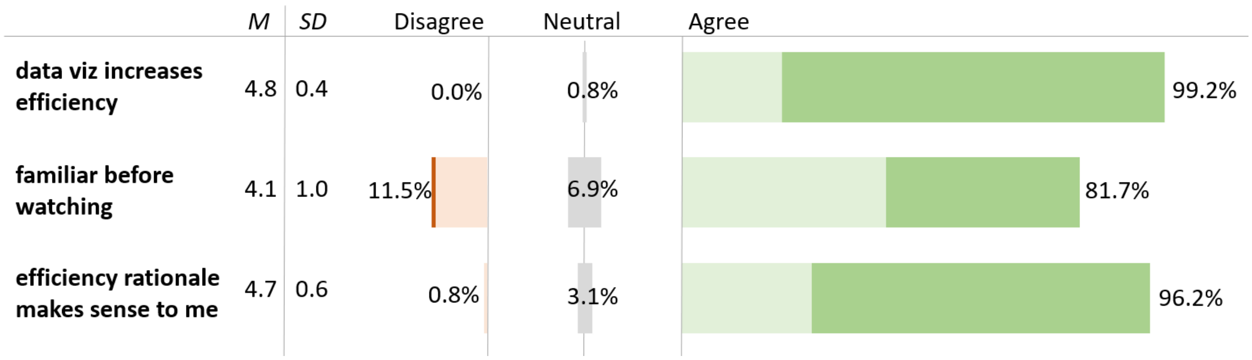
Note. Responses are grouped as low (*very hard* and *hard*) to the left in shades of orange, neutral responses in gray down the middle, and high responses (*easy* and *very easy*) to the right in green. *Very hard* and *very easy* responses are presented with more saturated colors in the bar. The numeric percent displayed outside the bar is the combination of low or high responses.

The Efficiency Rationale (RQ 1)

The first research question asked: *To what extent are participating evaluators (a) familiar with and (b) accepting of the increased efficiency rationale?* Figure 26 shows the extent of agreement with each question. Responses show that the vast majority of participants accept the "increased *efficiency* rationale" of data viz (*M* = 4.8; *SD* = 0.4) and think that the model "makes sense" to them (*M* = 4.7; *SD* = 0.6). A minority of participants (11.5%) indicated that they were not "familiar with" the increased *efficiency* rationale of data viz before watching the video (*M* = 4.1; *SD* = 1.0). Familiarity with this rationale before watching the video was positively correlated with participant data viz knowledge (r = .31, $p < .001$) but not with how frequently they use data viz.

Figure 26

Support for the Efficiency Rationale of Data Viz (N = 131)



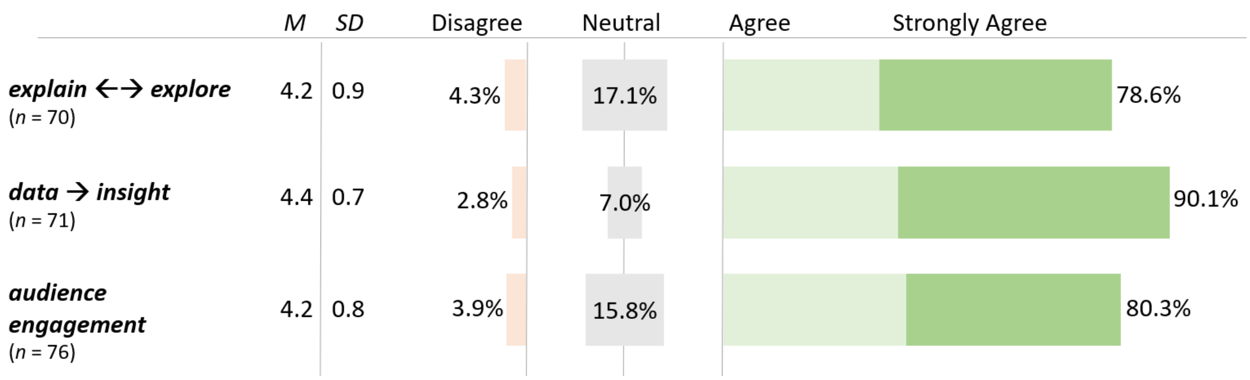
Note. Responses are grouped as negative (*strongly disagree* and *disagree*) to the left, neutral responses in gray down the middle, and positive responses (*agree* and *strongly agree*) to the right. *Strongly disagree* and *strongly agree* are presented with more saturated colors in the bar. The numeric percent displayed outside the bar is the combination of positive or negative responses.

The Efficiency Rationale (RQ 2)

The second research question asked: *To what extent do participating evaluators consider each model "adds value" to the efficiency rationale?* Answering this question relies on all participant responses to the question "If used, the [model] adds value to the 'efficiency rationale' for using data viz," regardless of how many model videos that participant had watched. Figure 27 shows that participants generally felt that all models add value to the *efficiency* rationale of data viz. While the responses shown here are not directly comparable to one another, ANOVA was conducted on the first responses to each model and is presented in Figure 28. The results are similar and demonstrate that there are no statistical differences between models on this question.

Figure 27

The [model] Adds Value to the 'Efficiency Rationale' for Using Data Viz (Regardless of View Order)



Note. Responses are grouped as negative (*strongly disagree* and *disagree*) to the left, neutral responses in gray down the middle, and positive responses (*agree* and *strongly agree*) to the right. The numeric percent displayed outside the bar is the combination of positive or negative responses.

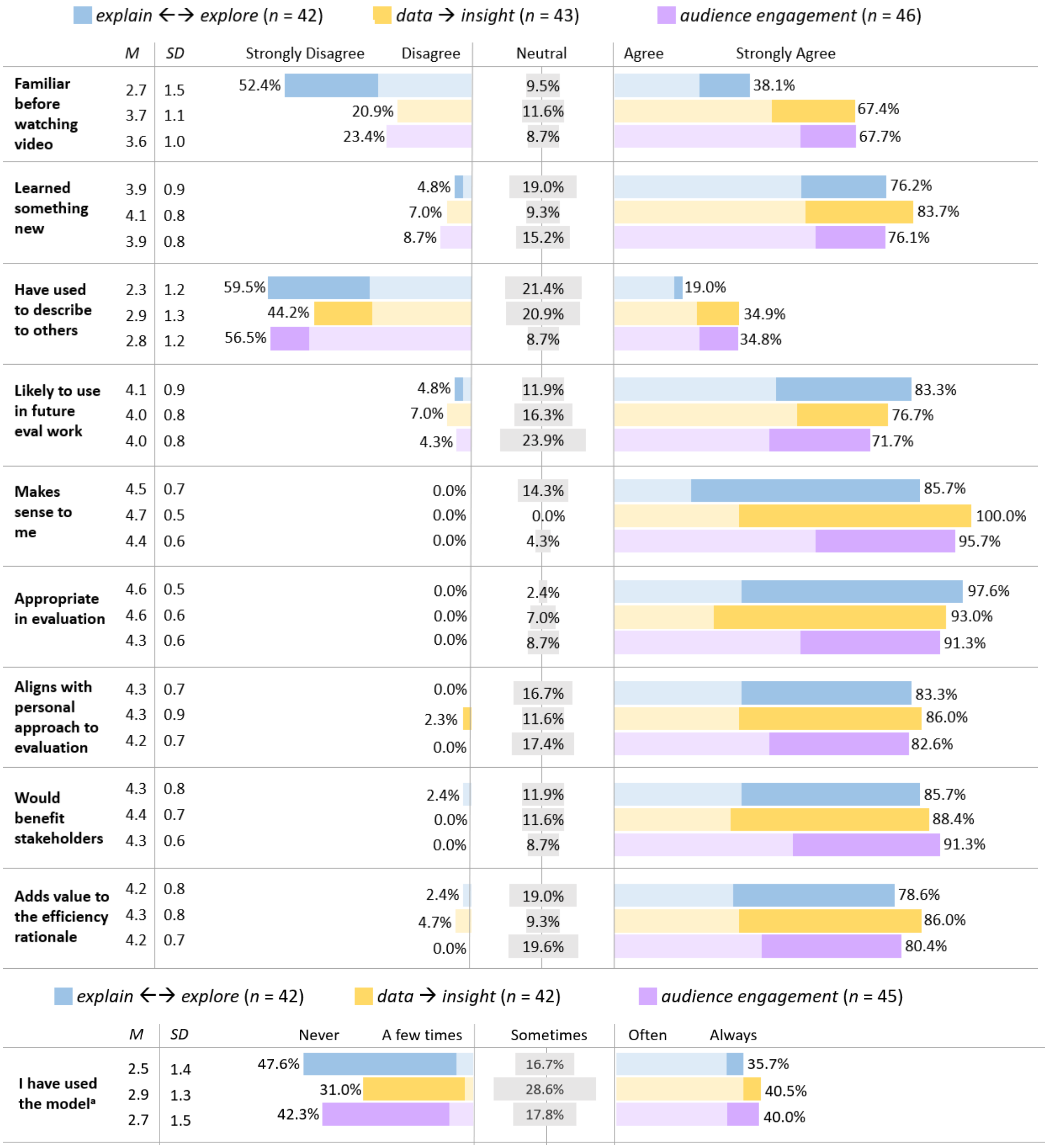
Directly Comparing the Models (RQ 3)

The third research question asked: *Are there significant differences between the models in evaluator (a) familiarity, (b) acceptance, (c) perceived appropriateness in evaluation, and/or (d) perceived usefulness in evaluation?* To answer this question, only the responses on the first model viewed by each participant were compared. Results are presented visually as well as statistically.

Figure 28 presents participant responses to the ten model questions grouped as negative (*strongly disagree* and *disagree*) to the left, neutral responses in gray down the middle, and positive responses (*agree* and *strongly agree*) to the right. *Strongly disagree* and *strongly agree* are presented with more saturated colors in the bar. The numeric percent displayed outside the bar is the combination of positive or negative responses. For instance, 26.2% (n = 11) of participants *strongly disagreed* that they were "familiar with the *explain* ←→ *explore* model before watching the video" and another 26.2% (n = 11) of participants *disagreed* with the statement. The total of 52.4% is presented to the left of the negative left side of the bar.

Figure 28

Comparison of Participant Responses After Viewing One Model Video (N = 131)



Note. [a]The question regarding frequency of using the model is missing two observations (1 each from AE and DI) and was presented to on a separate scale.

A one-way Welch ANOVA was conducted to determine if responses to any model questions were different based on model watched. The data were cleaned and analyzed using the ANOVA assumptions detailed in Appendix O. Participants were divided into three groups based on which model they had watched first: EE (*n* = 42), DI (*n* = 43), and AE (*n* = 46). As presented in Table 6, only three questions show significant differences between groups.

Table 6

ANOVA Results on Model Questions

	Welch ANOVA			Games-Howell Post Hoc Test			
	Stat	df	*p*		Mean Diff	SE	*p*
familiar-before	6.59	2	.002**	EE - DI	-0.960	0.276	.002**
				EE - AE	-0.849	0.270	.007**
learned-new	0.67	2	.515				
described-others	2.63	2	.078				
likely-to-use	0.64	2	.528				
makes-sense	3.70	2	.029*	DI - AE	0.303	0.111	.021*
appropriate-eval	3.22	2	.045*	No significant differences			
aligns-approach	0.26	2	.775				
benefit-stakeholders	0.38	2	.687				
adds-value-efficiency	0.32	2	.725				
have-used	0.70	2	.498				

Note. * indicates $p < .05$; ** indicates $p < .01$

Participant familiarity with the model before watching the video was statistically different between groups, Welch's $F(2, 82.51) = 6.59$, $p = .002$. Before watching the videos, participants were less likely to be familiar with the EE model than either of the other models. Games-Howell post hoc analysis revealed that the mean decrease in familiarity between the EE and the DI model (-0.960) was statistically significant ($p = .002$), as was the decrease between the EE and AE model (-0.849, $p = .007$).

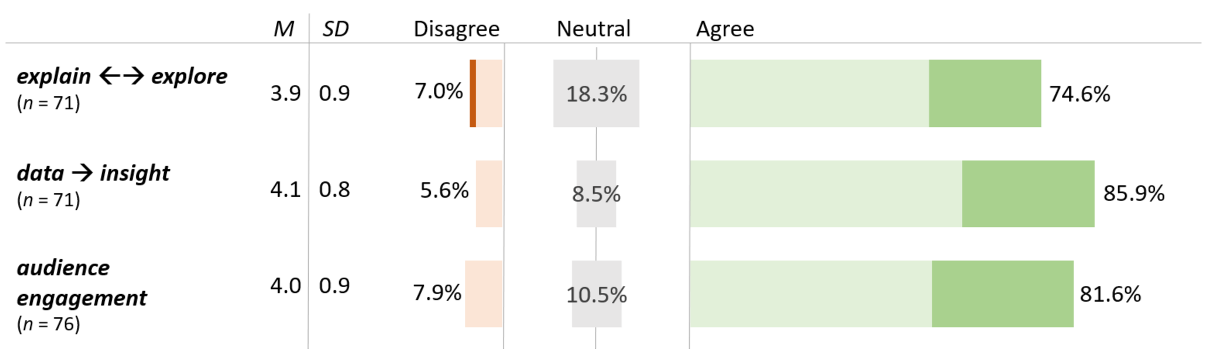
Participant agreement that the model “makes sense” to them was statistically different between groups, Welch's $F(2, 82.47) = 3.70$, $p = .029$. Games-Howell post hoc analysis revealed that the mean difference between the DI and AE models (0.303) was statistically significant ($p = .021$), indicating that participants felt that the DI model made more sense to them than the AE model.

The final question with statistical significance between groups, as demonstrated by Welch's $F(2, 85.13) = 3.22$, $p = .045$, was “The [model] way of thinking about data viz is appropriate in evaluation.” However, no significant differences were detectible when conducting post hoc tests with adjusted alpha. As seen in Figure 28, all models were rated very highly on this question, so it is unsurprising that differences were difficult to detect.

Research question 4 asked: *Did evaluators “learn something new” from watching the model videos?* While there is no statistical difference between models on the first model viewed (Welch's $F(2, 84.89) = 0.67$, $p = .515$), all models scored high on this question. Greater than 75% of participants responded *agree* or *strongly agree* that they “learned something new” about the first model they watched, see Figure 28. Viewing multiple videos does not seem to affect the perception that they learned something new. Combining all responses together, regardless of how many videos participants had watched, Figure 29 shows the similarly strong agreement that they had “learned something new.”

Figure 29

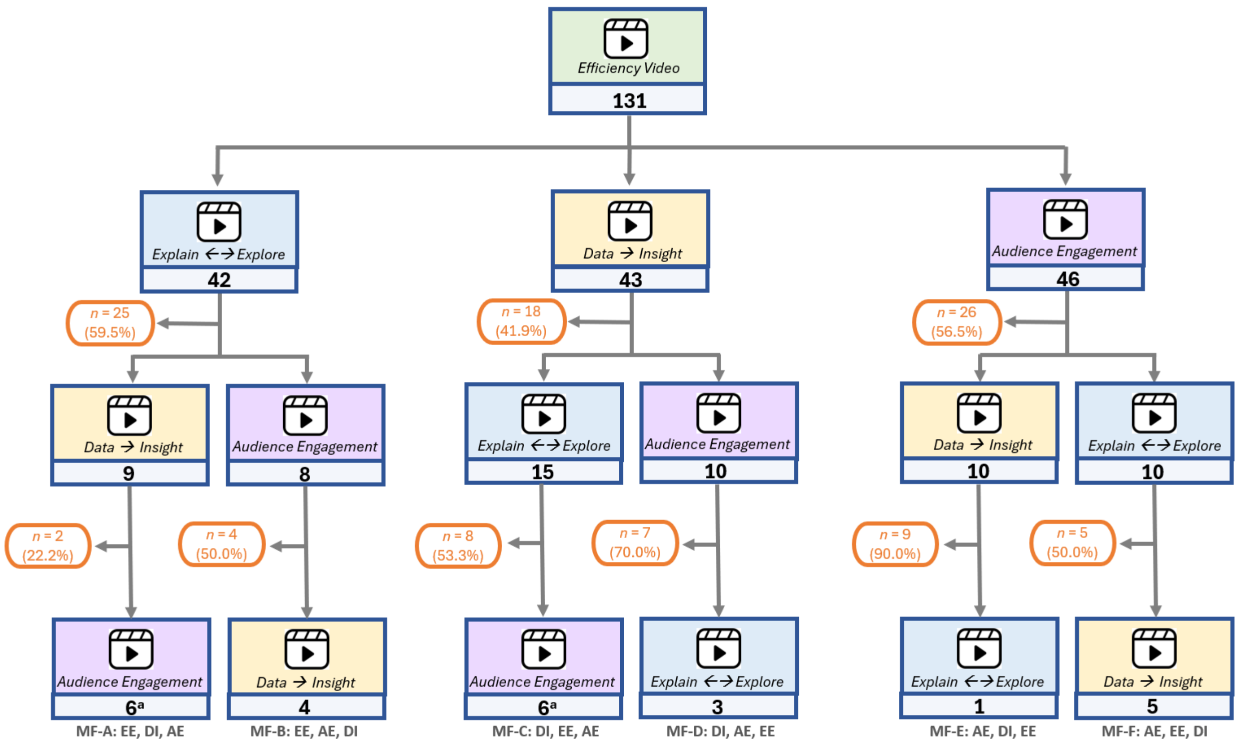
Learned Something New About the [Model] from Watching This Video (Regardless of View Order)



Research question five asked: *Are there significant differences between the model videos as demonstrated by (a) survey exit rates following a video or (b) evaluator model preferences after watching multiple videos?* This question offered another opportunity to directly compare the models. As each person was given the opportunity to leave the survey (exit) or continue and watch another video (continue), that decision provides a dichotomous variable for each video. The sample size met the chi-square requirement that all cells have expected counts greater than five after watching the first video, but some counts fell below that threshold for subsequent videos. Figure 30 shows the exit rate after watching each model video. While it appears visually that the exit rate is lower for the *data* → *insight* model, the exit rates are not significantly different ($X^2(2) = 3.08$, $p = 0.214$).

Figure 30

Exit Rates After Watching Model Videos (N = 131)

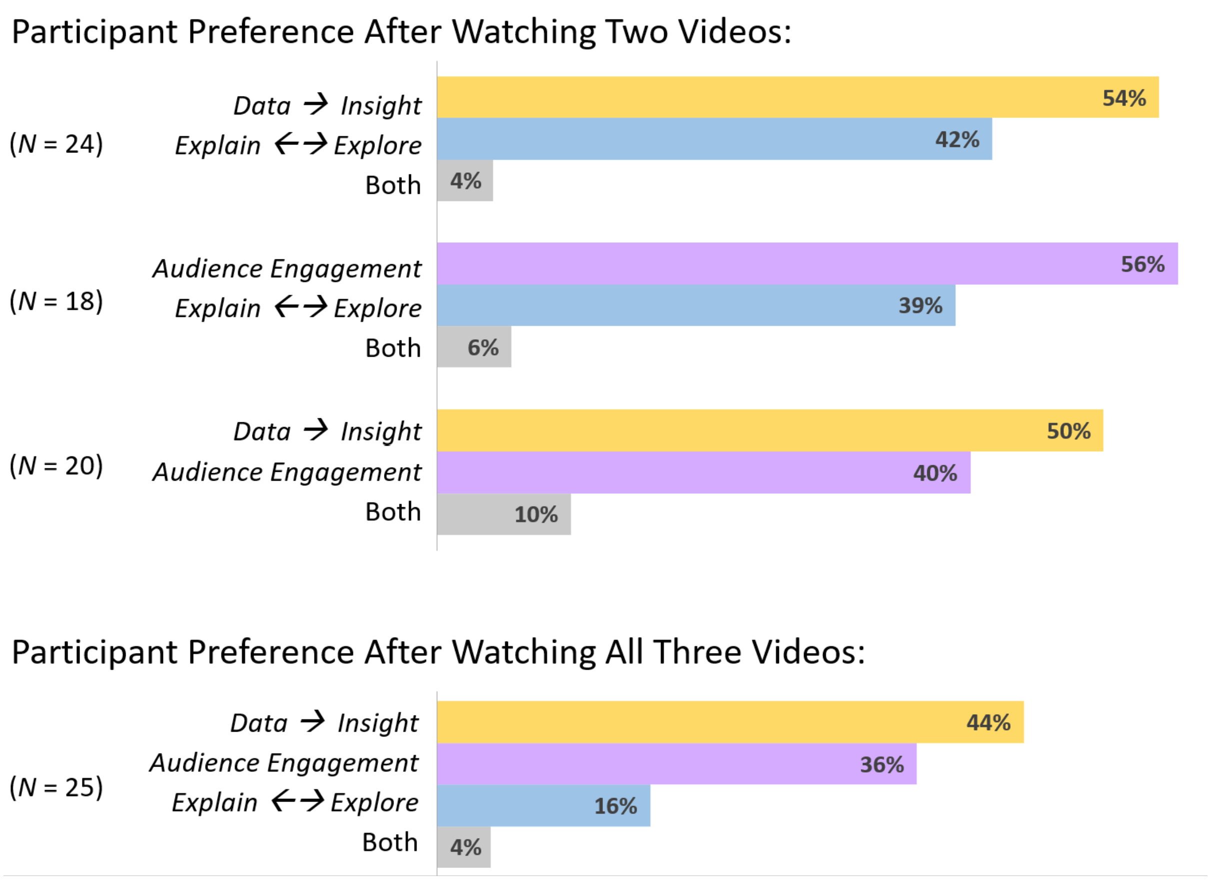


Note. [a]Count is reduced by $n = 1$ because one participant watched the video but did not supply any data for that video.

Although there are no statistical differences between exit rates, we can still consider model preferences after watching multiple videos. After watching two or more videos, each participant was asked which of the models they considered most useful to their evaluation work and why. Sixty-two participants (47.3%) stated their preference after watching two videos and 25 participants (19.1%) stated their preference after watching all three videos. Figure 31 shows these preferences. Although selecting both models was not an intended option, four participants skipped the question and then indicated that both models were valuable in the open-ended comment field. Those responses have been categorized as “both” in the visual, but were excluded from chi-square analysis. There is no difference in preference after watching two videos and even participant preference after watching all three videos is not statistically different ($X^2(2) = 3.25$, $p = 0.197$).

Figure 31

Participant Model Preferences After Watching Two or All Three Videos



Although there are no significant quantitative differences, participants also provided qualitative responses to the open-ended question, “Why did you chose the model that you did.” That qualitative analysis will be reported with research question six.

Suggested Model Improvements

The sixth research question asks: *Are model improvements suggested by this sample of evaluators?* This analysis was conducted by reviewing open-ended responses to the questions on perceived *strengths* and *weaknesses* from all models, regardless of watch order, as well as the open-ended comments provided after watching and directly comparing two or all three videos as discussed in the fifth research question.

Model Strengths and Weaknesses

Qualitative comments specific to each model's *strengths* and *weaknesses* were coded and themed using a conventional content analysis approach, where code categories were derived directly from the text data (Hsieh & Shannon, 2005). As I was focused on intended meaning within the context of the specific model, each model was considered separately such that codes and themes are not necessarily comparable across models. More details on this approach can be found in the Data Analysis section of this phase. Table 7 shows the number of qualitative responses provided for each question (i.e., the *explain* ←→ *explore* model received 44 comments regarding *strengths* and 36 comments regarding *weaknesses*) as well as the mean, minimum, and maximum word count for those responses. The "themes" column of Table 7 provides details on the themes identified as *strengths* and *weaknesses* for each model.

Table 7

Strength and Weakness Themes Identified Per Model

Responses *n*	Word Count Mean [min – max]	Themes
explain ←→ explore		
44	18.6 [3 - 72]	**Strengths:** * Useful for evaluator: Helps make viz decisions (*n* = 19) * Simplicity of the model (*n* = 7) * Emphasizes audience engagement (*n* = 6) * *Explore* role for the evaluator (*n* = 3) * Promotes use (*n* = 3)
36	38.0 [3 - 441]	**Weaknesses:** * Just a way of thinking, doesn't say "how" or "benefits" (*n* = 8) * Not as binary as appears, can be both (*n* = 7) * Does not address visualcy and comprehension (*n* = 4) * Over-simplifies complexity (*n* = 3)
data → insight		
51	17.5 [3 - 61]	**Strengths:** * Clear, iterative, directional, phases, steps, continuum (*n* = 11) * Intuitive (easy to understand) for evaluators and/or stakeholders (*n* = 11) * Argument for data viz efficiency (*n* = 8) * Direct connection to evaluation processes (*n* = 7) * Connection to cognitive processes (*n* = 5)
39	25.7 [5 - 76]	**Weaknesses:** * Too linear or too simple (*n* = 6) * Just a way of thinking, not clear how it can be used (*n* = 3)
audience engagement		

52	19.1 [3 - 59]	**Strengths:**
		* Encourages use/decisions/action (*n* = 13)
		* Useful as a mental model (*n* = 10)
		* Emphasis on a specific aspect of the model:
		   * Increases or prioritizes connection/engagement (*n* = 8)
		   * ECB/learning/knowledge/understanding/memory (*n* = 5)
		   * Professionalism/confidence/credibility component (*n* = 3)
		* Specific to eval (*n* = 4)

48	28.5 [4 - 286]	**Weaknesses:** * Concerns related to audience (difficult/time consuming) (*n* = 10) * Data viz is hard to do (*n* = 6) or data viz has weaknesses (*n* = 4) * Concerns over links between professionalism and credibility (*n* = 5)

Explain ←→ Explore Model. The most often mentioned *strength* of this model is its usefulness to the evaluator (43.2%) in helping to clarify their intent with the data viz and determine how best to present it. Another frequently mentioned strength is the sheer simplicity (15.9%) of the model – for both the evaluator and for others. With only "two directions (arrows) to look [at] and think about," it is considered an easy way for evaluators to think about their reason for creating a viz and "stakeholders are familiar with the terms *explain* and *explore* over other terms we frequently use." Consistent with the overlap between the models, comments also suggested the value of using the *explain* ←→ *explore* model for audience engagement (13.6%) and there were numerous mentions of words associated with other models such as "sense-making," "decision-making," and "insight."

Three participants specifically commented on the value of the model in articulating the benefits of data viz to evaluators in their own work. One person indicated, "it helps justify the use of exploratory data viz, even if it's just for the analyst, in a way that I haven't seen before" and another said that it "remind[s] evaluators that they can use data viz for their own exploration of the data."

The most often mentioned weakness of the model was that it is only a "a way of thinking" and does not "provide any guidance to those creating the data viz" nor "articulate a specific benefit" (22.2%). Half (n = 4) of the participants mentioning this weakness specifically brought up the fact that the model does not address the visualcy and comprehension needs of the audience.

The next most often mentioned *weakness* with the model was with the representation of *explain* and *explore* as discrete categories (19.4%). While these categories may make sense within the model framing from "the perspective of the person having the experience," the comments suggest that evaluators are interested in a more nuanced model with comments such as, "sometimes these two things (explain/explore) happen in tandem and are not binary," "they need to complement each other," and "I think people would get more value by exploring what something along the spectrum looks like." Relatedly, one participant clearly articulated that a continuum is not the correct visual for this model:

> By using these as two ends of a "continuum," it confuses intention with time and also becomes confounded with "state". You must "know" or be aware of something, before attempting to

explain it; You explore (curiously) to find something else to know. There is no "opposite" here; only two of possibly many distinct purposes.

There were a number of comments that could not be themed into a weakness of the model as they were from participants reflecting on the drawbacks of pursuing activities mentioned in the model, not the model itself. Of note, three people wrote about the time and effort needed to conduct explore data viz with comments such as, “It is easy to find yourself taking too long in the exploration phase, without necessarily finding a good explanation method to use in the end.”

It is notable, but not surprising, that one of the model’s greatest strengths – its simplicity – is also viewed as a weakness. Multiple participants commented on the inherent problems in over-simplification – both in reducing data into a visualization and in reducing complex motivations and behaviors into a model. Overall, the *explain* ←→ *explore* model is a simple, easy to explain, and easy to use model that doesn’t provide enough prescriptive direction and may be a little too simple and reductionist.

Data → Insight Model. In total, 43.1% of comments regarding *strengths* of the *data* → *insight* model related to the model being clear or intuitive. I carefully reviewed these comments multiple times and believe that they fall into two, evenly split, and related sub-themes. Evaluators thought that the model was clear (n = 11) because the model itself is presented in an “iterative,” “linear,” “directional,” “progression,” or “continuum” with “phases” and “steps” that participants found easy to understand. They pointed out that the organization of the model “helps you think about why you're doing your data viz and therefore what will work best” and that the linearity can make it more “clear about what phase you're in.” There were also comments suggesting that it is “always useful to name the components of a theoretical construct” and even that the phases of the model could be converted into a “checklist.” The other half of these comments (n = 11) were themed into the model being “intuitive” because it was “easy to grasp” and easy for evaluators and/or stakeholders to understand. Multiple comments suggested that the model is “adaptable to many situations” and that the model’s “applicability to fields outside evaluation … will connect with stakeholders.” Five additional comments (9.8%) specifically mention that emphasizing “cognitive processes” is a *strength* of the model and suggested that the model “reflects the scientific method” and “aligns with how humans think.”

Another identified theme is that the *data* → *insight* model is an “argument for data viz efficiency” (15.7%). Comments suggest that the model provides “a theoretical basis for the need and value to use data viz,” can “help evaluators understand why data visualization can be so important,” and even “justifies the need and time to use data viz to help interpret and share results.”

Comments also suggested that the *data* → *insight* model shows a direct connection to evaluation processes (13.7%) with comments such as “it is linked directly to the purpose of program evaluation generally which is all about providing insight into the how and why of a program” and that the model “makes it clear that we are not just collecting information for the sake of having data, but because we want to build upon the data to get to the point of insight, where data can be used for decision making.”

The most often mentioned *weaknesses* was that the model is “too linear” or “too simple” (15.4%) and a further 3 people suggested that the model was “just a way of thinking and that it isn’t clear how it can be used.” There were fewer *weaknesses* themed from these model comments than for the other models. There were many valuable criticisms provided about the model videos such as recommending less “jargon” in the explanation, indicating that the video was too “boring and academic,” or suggesting that it should be “distill[ed] to just a few (1-3) key points.” However, other comments could be themed as *weaknesses* because they were more appropriately themed as strengths (e.g., “I think it’s very straight-forward”) or were about concerns related to data itself, such as pointing out that “data integrity are critical - without solid measurement, it is simply marketing” and difficult to theme to the model.

Overall, the *data* → *insight* model is considered clear and intuitive, applicable to evaluation, and a benefit to evaluators and stakeholders, but it could benefit from edits to the video explanation.

Audience Engagement Model. Of the 52 comments about the *strengths* of the *audience engagement* model, 25.0% specifically tied the model to encouraging use, decision-making, and/or taking action with data. Connection to a *U-FE* approach was clear in comments such as, “I follow a utilization-focused evaluation framework … the audience engagement model provides me another framework to accomplish a similar goal.” Comments indicated that the model “goes beyond ease of understanding to encourage application” and suggested specific ways that the model accomplishes this such as “promot[ing] dialogue,” “distill[ing] key messages,” or “invit[ing] buy-in more quickly.”

Another 19.2% of comments spoke directly to the value of the model as a mental model with comments such as, “I think it helps to concretely describe processes that already happen with more clarity.” They suggested benefits of using this mental model such as “focusing attention,” “mak[ing] you think more critically about the audience and their needs,” and “emphasiz[ing] concrete outcomes.”

A number of comments are listed as themes in the table but are not discussed here because they reiterated portions of the model the participant found interesting or beneficial as strengths – rather than discussing the model as a mental model. While not a large number, 4 (7.7%) of the comments suggested that the model supports evaluation processes such as “it shows [the audience] that evaluation is valuable (not scary or punitive) and can be used to ask other questions of interest to the audience” and “it takes into account the reality of evaluation as an industry, not just a science.”

The most commonly mentioned weaknesses of the model were not about the model itself, but were concerns related to the practicality of engaging an audience (20.8%). These concerns discussed how difficult and time consuming it is to engage the many different groups needed as well as concerns that “clients don't always WANT to be engaged,” with one evaluator stating, “not everyone is as excited about evaluation as evaluators.”

The next most common theme was also not about the model itself, but that data viz is difficult (12.5%) and data viz itself has weaknesses (8.3%). Comments emphasized the difficulty in "becoming fluent in data visualization," and the "additional resources needed" to hire or acquire expertise in design principles. Relatedly, comments focused on problems inherent to data viz itself such as over-simplifying data or misleading an audience.

There were a number of individual comments suggesting specific improvements to the model, but these are not themed nor discussed here. However, 10.4% of the comments specifically mentioned concerns over the model's linkage between professionalism and credibility with data viz. Comments suggested that the "professionalism piece" is over emphasized and others questioned the role of data viz in the relationship because there are "other factors at play" or even that data viz can hurt credibility with one person saying, "in some of my evaluations, the use of data viz was seen as less scientific and credible, and the stakeholders preferred text." Others indicated that data viz or even engagement itself should not be linked to professionalism. There were also some comments that appeared to suggest potential negative downsides of engagement itself such as, "I wouldn't want to see an evaluator trying too hard to please an audience just to generate additional contracts." It was impossible to pursue these thoughts from the data, but this could be investigated further.

Overall, comments suggest that the *audience engagement* model is a useful way of thinking about data viz in evaluation because it clarifies and describes processes that already happen and reflects the "reality of evaluation as an industry, not just a science." However, there are concerns that the model over emphasizes the importance of data viz and/or the importance of engagement.

Model Preferences After Watching Multiple Videos

After watching two or more videos, each participant was asked which of the models they considered most useful to their evaluation work and why. Sixty-two participants (47.3%) stated their preference after watching two videos and 25 participants (19.1%) stated their preference after watching all three videos. Figure 31 in research question five shows these preferences. Table 8 provides a summary of the qualitative comments regarding which model was preferred and why.

Table 8

Rationale for Choosing One Model Over Another After Watching More Than One Model Video

Models Compared	*N*	Word Count Mean [Min - Max]	Summary Quote	Summary
explain ←→ explore *data → insight*	20	24.6 [10 - 63]	"I think they are both useful...& even more so in tandem."	There were several suggestions that the DI model is incorporated into the EE model, but there was also a comment suggesting the opposite nesting relationship.
explain ←→ explore *audience engagement*	16	30.4 [4 - 76]	"They can both be useful and really are not mutually exclusive."	EE was seen as easier to understand and explain, but AE was seen as more aligned to eval, more relevant to eval work, and more useful as a way of thinking about audience and purpose.
data → insight *audience engagement*	16	26.3 [3 - 78]	"Really it's a toss up. Both are sensible and useful."	Model preferences used the same rationale regardless of model (i.e., each model encourages use more than the other, leads to insight more than the other, etc.)
All three models	19	32.7 [8 - 77]	"Considering that they can all happen in one session, I think they're equally useful."	AE has the most information for an evaluator and encompasses the other two. DI is based in cognitive science and is easy and intuitive to apply. EE is simple, maybe too simple.

Comparing Two Models. It was previously noted that four participants skipped selecting one model over the others and indicated that both models were valuable in the open-ended comment field. Upon reviewing the qualitative comments, even those participants who did select a model often went on to explain that both models are useful and/or complementary. Both models being complementary and serving different purposes at different times was the leading theme regardless of which two models were compared. No new themes of model strengths arose here; the same themes of model preferences that were discussed when analyzing each model independently arose in this analysis. Participants explained that their preferred model was simpler, more intuitive, etc., based on their experience.

Comparing All Three Models. The suggestion that the models are connected is prevalent in the comments, as befitting the already well accepted overlap between the models. Comments suggested that the *audience engagement* model "encompasses the other two" and "has the most information in the model, and there is more there to delve into." The *data* → *insight* model was described as "the most foundational of the three, like the other two are extensions off of that basic premise" and was also described as "the most directly applicable to everyday use and work." Both *explain* ←→ *explore* and *data* → *insight* were described as "simplistic," "fairly obvious," and "more conceptual" when compared to the *audience engagement* model. While terms such as "umbrella concept," "encompasses," "foundational," etc. were used, the presumed hierarchy of the relationship(s) cannot be determined with any confidence from this limited data set.

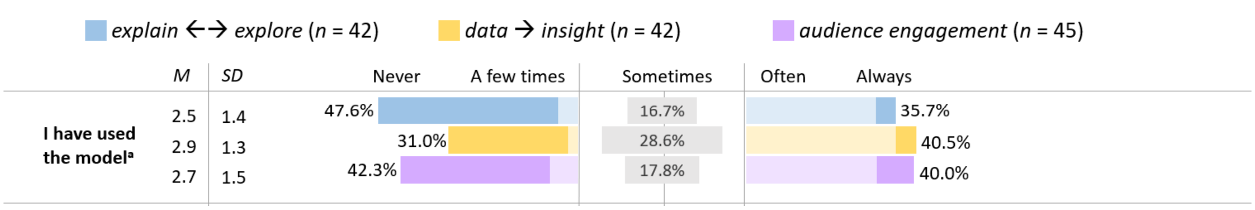
Experience Using the Conceptual Model in Evaluation

The final research question considers evaluator experience using each conceptual model in their evaluation work: *For each model under consideration, (a) to what extent and (b) in what ways, have participating evaluators used the model, and (c) what outcomes do they attribute to that use?*

Figure 32 shows that many participants have *never* used the model they were presented first in their evaluation work (EE = 42.9%, DI = 28.6%, and AE = 37.0%).

Figure 32

Comparison of Participant Responses on Use After Viewing One Model Video (N = 131)



Note. [a] This question is missing two observations (1 each from AE and DI).

While there is no statistical difference between models regarding the extent to which evaluators report using these models in their evaluation work, reviewing responses from all models – regardless of watch order – provides more detail on the ways they have used each model and the outcomes they attribute to that use. Table 9 provides details on the qualitative comments analyzed and the major themes identified in each model.

Table 9

Use Themes Identified Per Model

	Frequency		Word Count		
Response	*n*	%	*n*	Mean	Themes
	explain ←→ explore				
Always	6	8.5	6	45.5	* Evaluator uses "explore" as an "analyst" (*n* = 11)
Often	20	28.2	14	33.3	* Participation increases ECB, engagement, and/or use (*n* = 6)
Some	9	12.7	4	21.8	* Familiar with concept, but not named model (*n* = 3)
Few	6	8.5	5	36	
Never	28	39.4	1	21	
Missing	2	2.8	0	0	
Total	**71**		**30**		
	data → insight				
Always	7	9.9	5	42.2	* Use the principles, but not the named model (*n* = 6)
Often	19	26.8	15	24.2	* Better stakeholder understanding (*n* = 5)
Some	19	26.8	10	25.4	* Helps to see patterns (*n* = 5)
Few	3	4.2	2	42.5	* Fosters stakeholder engagement or participation (*n* = 5)
Never	21	29.6	0	0	* Using viz of qual process (Ex: theory of change, contract
Missing	2	2.8	1	42	negotiations) (*n* = 3)
Total	**71**		**33**		
	audience engagement				
Always	10	13.5	9	22.1	* Better understanding (*n* = 9)
Often	20	27.0	16	30.6	* Appreciate evaluation/evaluator (*n* = 7)
Some	12	16.2	8	48.5	* Use the principles, but not the named model (*n* = 6)
Few	5	6.8	3	27.7	
Never	27	36.5	2	50.0	
Missing	2	2.7	1	27.0	
Total	**74**		**40**		

Note. Number of responses and word counts are associated with the response categories “always” to “never” as an indication of how many participants in each response category provided data. Themes are drawn from all responses.

Explain ←→ Explore Model

The role of data viz to *explain* findings in reports to stakeholders was present in the data, as expected. Also, as expected, several participants acknowledged using the model in practice without ever having called it by a name because the concept is so "natural." A more interesting finding was that, after watching a video describing the *explore* role of data viz, evaluators discussed data viz from this perspective to a much greater extent than I uncovered in my literature review or in previous phases of this study. Of the 30 open-ended comments received from the 71 participants who viewed the *explain* ←→ *explore* video, 33.3% (n = 11) specifically mentioned that they use data viz as a way to *explore* data with comments such as, "I always apply some [data viz] time for myself when analyzing the data to uncover interesting findings. Sometimes these make it into a presentation or report to discuss with stakeholders." They suggested that data viz to *explore* was the evaluator in an "analyst role" and that *exploring* the data visually led to "stronger final versions" and "clearer reporting and greater understanding."

There were also a number of comments (n = 6; 20.0%) related to the value of using *explore* visuals collaboratively with stakeholders. Participants specifically mentioned ECB and suggested that participatory data viz leads to "active participation" and "greater use." Two participants framed *exploring* data with clients as an opportunity to achieve additional insights due to the client's "better content knowledge" or additional questions after interacting with the data. One interesting comment on the value of sharing exploratory visuals with clients was, "organizational leaders have appreciated exploratory visuals because they're not a 'report' or judgmental, but just a pattern to question or look into."

Data → Insight Model

Unsurprisingly, a number of participants (n = 6; 18.2%) commented that they have used the model or aspects of the model without having a name for the model: "I have often used the approach of moving from data to information to knowledge to insight for reports and presentations, but I have not intentionally used the data to insight model and made it explicit.” This is consistent with previous findings and the literature review. What was more interesting in this analysis was the number of people who discussed the model with terms I associate with the *efficiency* rationale, such as the model’s contribution to “better understanding” (*n* = 5; 15.2%) or “seeing patterns” (*n* = 5; 15.2%). This may suggest that participants consider the *data* → *insight* model as a foundational or underlying model. An alternative explanation is that this analysis considers responses after evaluators had watched multiple videos. As there is tremendous conceptual overlap between the models, this is not unexpected and they did use terminology in these responses that were described in other videos, such as *engaging* stakeholders (*n* = 5; 15.2%) or *exploring* data.

Audience Engagement Model

As with the other models, many of the comments were simply reiterations of the portions of the model that the participant finds valuable or a summary of the model. The most commonly discussed result from using the *audience engagement* model concerned increased understanding (22.5%) on the part of both the client and the evaluator. Comments suggested that this increased understanding occurs due to deeper and more involved questions. The next most commonly discussed result emphasized audience appreciation for the evaluator or evaluation itself (71.5%). Examples provided were consistent with previous discussions on this model: clients have “asked for more data visualizations,” have “asked us to replicate something similar for them,” and have become “repeat customers.”

Consistent with the other models, another commonly mentioned theme was that participants have used the principles without having a named model (15.0%): “I have not used this model formally, but the concepts included are often concepts that I use to train and grow less experienced evaluators in their practice.”

Phase III: Discussion

The non-significant quantitative findings combined with the qualitative comments make it abundantly clear that the models overlap and are complementary. While a few research questions sought to directly compare the models, this exploratory research was not intended to provide definitive relationships or hierarchies between the models. Rather than comparing the models, this discussion will emphasize that all models are considered appropriate and useful in evaluation and all models offer potential avenues for continuing education and professional development. One benefit of conceptualizing additional reasons for using data viz is that it may provide evaluators reasons to "justify" the time, skill, and effort required to use data viz. Evaluators feeling a need to "justify" using data viz, along with the finding that they are interested in using data viz to *explore* data in the analyst role, suggests that there may be more interest in reasons for using data viz than the efficiency reason commonly discussed within evaluation.

However, data viz is not a panacea. Evaluators are interested in reasons for using data viz beyond efficiency, but they have realistic concerns over potential downsides of data viz such as ethical concerns related to focusing on client satisfaction and securing contracts over findings, the potential to emphasize findings based on ease of visualizing them rather than importance, and a host of underlying concerns over data quality and validity that exist regardless of data viz.

All Models are Appropriate and Useful in Evaluation

All three models scored 4.0 (*agree*) or above on all questions related to their appropriateness and usefulness in evaluation (i.e., *likely to use in future eval work*, *appropriate in eval*, and *would benefit stakeholders*). The sense that no models are "bad" was apparent in the qualitative comments, as well. Particularly when directly comparing models, participants again displayed a uniformly positive response to the models and sought to nest them or combine them in their comments. However, suggesting an integrated model is outside the scope of this dissertation and may not be desirable. The flexibility participants displayed in applying the models to their work suggests that an integrated model may not have more value than simply articulating these, and possibly other, models to evaluators and allowing them to decide which portions benefit them and when.

Frameworks Have Value and Offer Avenues for Engagement and Training

There was a prevailing sense in Phase III that simply defining each model as a mental framework has value. When used this way, conceptual frameworks are not intended to be rigid structures; evaluators can combine and integrate the portion(s) that are useful to them at any given time. In addition to finding all three models appropriate and useful, a resounding theme throughout this analysis is that participants had used the model or aspects of the model without having a name for it. Once terminology was associated with the model, they were able to critique it and suggest ways to build upon it. Importantly, comments used the exact same rationale to support different models (i.e., each model encourages *use* more than the other, each model leads to *insight* more than the other, etc.). I suggest that there is a benefit simply in naming and defining the models so that evaluators can consider using aspects of each in their work.

It was clear from the comments that evaluators already use key aspects of the models even if they don't use the same terminology. One comment on the *audience engagement* model, "I think it helps to concretely describe processes that already happen with more clarity," in essence summarizes a key reason for articulating a conceptual model. Yes, many experienced evaluators already understand and use portions of the model relevant to their work, but described frameworks provide opportunities to easily teach key concepts to less experienced evaluators. While there were a number of comments suggesting that the models hold value for newer evaluators and in teaching or learning about evaluation, support for the reflective and teaching aspects of the models are not limited to new evaluators. An evaluator with ten years of experience said of the *audience engagement* model: "There is a lot there to digest. I want a print-out to think about all the ways it applies to my work, and other ways I can make it work for me. There is more to consider than a quick video."

Also, in this sample, there was no correlation between evaluation knowledge and data viz knowledge, so it is likely that even experienced evaluators can benefit from learning more about data viz. Given the lack of relationship between evaluation knowledge and data viz knowledge and an understanding that evaluators enter the profession from a variety of fields and with a wide variety of preparation (e.g., it cannot be assumed they have the same baseline knowledge), I suggest that all three models could be provided to evaluators, regardless of their experience.

There is Continued Need and Desire for Data Viz Training and Tools in Evaluation

Although the only incentive for participating in this study was an opportunity to learn something about data viz, 131 professional evaluators voluntarily spent their time watching the videos and contributing their knowledge. This speaks both to their support of ongoing research in their field and to their interest in data viz. They were only asked to watch one model video, yet nearly half (47.3%) voluntarily watched two and 19.1% watched all three videos. This suggests that there is evaluator interest in learning more about data viz *use*. Phase III findings also suggest that there is a need for more data viz training and education based on the discrepancy evaluators reported between their knowledge and skills in evaluation (M = 3.9; SD = .07) and in data viz (M = 3.4; SD = .07) and on their heavy use of data viz in their evaluation work. Findings showed that participants *usually* (38.2%) or *always* (42.7%) use data viz in their evaluation work (M = 4.2; SD = 0.8) but *learning* new ways of displaying data is an infrequent activity with the majority of participants (87.1%) doing so *once a month or less*.

While there are data viz textbooks, conference workshops, and fee-based courses available to evaluators, this study suggests that there may also be unmet interest and need in smaller, briefer offerings such as these videos. Findings also suggest that evaluators may be interested in more abstract learning opportunities that emphasize ways of thinking. AEA has offered data viz content in their “coffee-break-length demonstrations of 20-minutes” in the past and should continue to do so. As they develop and provide learning opportunities, professional evaluation groups, educational institutions, and content creators should consider that the topics covered in this study (e.g., data viz, cognitive science conceptual models, audience engagement, etc.) are of interest to evaluators and that the format they were offered in (e.g., brief explainer videos) were appreciated.

Emphasis on “Justifying” Data Viz

Beyond finding the models appropriate and useful in evaluation, a number of comments suggested that considering these multiple reasons for using data viz that go beyond efficiency “justify” the time, skill, and effort required to use data viz.

It is my unquantifiable observation from this data that some evaluators feel data viz may not be worth the effort it takes to do it well. One person stated this eloquently: “The efficiency model is problematic because it defines efficiency from the audience perspective; but from the creator's perspective, data viz is very often NOT efficient - it can be quite resource intensive.” Other participants commented on the time and money needed to hire or acquire expertise in this area. As they expressed an interest in the models and enthusiasm over having additional reasons to use data viz in their work, it appears that they want to have a reason to justify spending their time and money on data viz activities.

As this sample includes only participants who are interested enough in data viz to complete the survey, there might be an even greater reluctance to spend time and money on data viz from the general population of evaluators. But, there may also be this same latent interest or curiosity over data viz such that an understanding of additional reasons for using data viz could result in more data viz use. If evaluators should become competent in data viz, on top of all the other aspects of evaluation, there is even greater value in providing them with learning opportunities and multiple rationales for data viz activities. As busy professionals with many competing needs, it is understandable that evaluators may prefer activities that accomplish multiple goals simultaneously. Conceptualizing uses of data viz that extend beyond the *efficiency rationale* and support evaluation in other ways may encourage these evaluators to engage in data viz. Of course, this concept is not limited to data viz; it suggests that conceptualizing more reasons to engage in desirable but effortful activities can benefit evaluation by encouraging these activities.

Evaluators Use Data Viz to Explore Data

Using data viz to explore data is an activity that I suspected evaluators were performing, but not discussing. As stated before, there are more references to *explain* than *explore* in evaluation literature and in the interviews analyzed in this dissertation. I had questioned whether evaluators might be conducting *explore* activities and simply not discussing them, or perhaps they are conducting them without conceptualizing those as “data viz,” or both.

Once presented with the heavy emphasis on *explore* in the *explain* ←→ *explore* model video, comments supported that evaluators do *explore* data visually and put themselves in the “analyst role” to a much greater extent than I uncovered previously. While this is not a surprise finding, it does suggest that evaluators are using data viz for their own purposes and to *explore* data without necessarily labeling it as such. Further, they find value in these activities. This, in turn, suggests that applying the label “data viz” to *explore* activities and expanding the emphasis beyond reporting provides additional opportunities to incorporate data viz into evaluation. This conceptualization of data viz to support the evaluator’s process, rather than the audience focus so clearly apparent in evaluation, may require additional training and support within the profession.

Audience Engagement Model Concerns

Unsurprisingly, data viz is not a panacea. Evaluators are interested in reasons for using data viz beyond efficiency, but they have realistic concerns over potential downsides of data viz such as ethical concerns related to focusing on contracts and client satisfaction over findings, the potential to emphasize findings based on ease of visualizing them rather than importance, and a host of underlying concerns over data quality and validity that exist regardless of data viz.

Overall, there were not very many concerns raised regarding the models, but the weaknesses that were discussed often applied to one of these areas and were discussed more often in association with the *audience engagement* model. One quote presented as a weakness of the *audience engagement* model exemplifies this:

> It can focus attention on those things that are easier to show visually. Complex results that don't show clear findings can be overlooked, and qualitative findings also sometimes can be missed because they are not as visual and can take more time to understand.

While this concern is reminiscent of the adage "you get what you measure" and is a potential weakness of data and data visualization in general, it was not closely associated with the other models. For the *audience engagement* model only, comments that data viz is difficult (12.5%) and data viz itself has weaknesses (8.3%) rose to the level of themes. These concerns may be exacerbated by the *audience engagement* model's emphasis on using data viz to engage, thereby preferencing visually appealing but potentially less important data. This provides a timely reminder that evaluators should focus on the purpose and intent behind the data viz and remain focused on evaluation questions and substance. The activities described in the *audience engagement* model should support the evaluation.

Phase III: Limitations

Insufficient Information on Time Spent Watching Videos

During the survey pilot, several people suggested that I slow down the presentation speed of the videos. One pilot participant stated, "I paused the video several times and re-watched to make sure I was understanding." They were able to do this because I had enabled the YouTube video controls in the survey software to allow users to adjust the speed for their comprehension. Hoping to analyze the time spent watching each video and determine whether some videos took longer to process, I added a timer on each video page in Qualtrics. However, the questions were on the same page as their corresponding video and participants were able to leave and return to the survey, so the timer data was useless for analysis. This was a missed opportunity. A better understanding of whether videos were slowed down (or sped up) or watched multiple times could have added a better understanding of the perceived conceptual difficulty of each model.

Video Content, Length, and Animation Skill Could Create Confounds

There was nearly a minute difference in video length between the shortest model video (*explain* ←→ *explore* at 3:04) and longest video (*data* → *insight* at 3:57). Significantly different lengths introduces an obvious confound when attempting to compare the videos. Furthermore, my animation skill developed while creating these videos such that the final videos display varied levels of editing skill, animation ability, and sound quality. Videos were edited and re-recorded to attempt to mitigate this, but confounding differences no doubt exist.

However, the area that provides the most opportunity to significantly affect this study is in the content of the videos. I necessarily had to prioritize some concepts over others to condense such involved conceptual models into videos under four minutes. I suspect that this led to the surprise finding in Phase III: Participants were significantly less likely to be familiar with the *explain* ←→ *explore* model than either of the other models. There were very few quantitative differences between the models and this – the strongest quantitative finding – is in stark contrast to my hypothesis. When proposing this project, I hypothesized that evaluators would be the most familiar with the *explain* ←→ *explore* model because it is mentioned frequently in evaluation. Their data showed the opposite.

This puzzling finding is likely due to my emphasis on the *explore* portion of the model in the video I created. I also emphasized that the anchors are from the *perspective of the person having the experience*. This particular framing is provided by Kirk (2016), a resource not specific to evaluation. The result was a video that was more in-depth and abstract than the simple words "explain" and "explore" imply. Based on qualitative comments, which support that participants have used the model without naming it, and the prevalence of using data viz to *explain* findings in evaluation, I suspect that the low familiarity recorded for the *explain* ←→ *explore* model refers to complexity I introduced at the *explore* end of the continuum.

An Even Larger Sample Size Could Have Provided Significant Findings

The goal of Phase III was to collect quantitative data with a larger, representative sample of AEA members in the hopes that findings would be generalizable to AEA members. The sample size was larger than the minimum required to conduct the chi-square tests used and the sample was representative of AEA membership. Unfortunately, if the apparent differences are true differences, the sample size is insufficient to detect those differences using chi-square tests. In fact, the same percentages of participants leaving or continuing the study would have been statistically significant with a larger sample size.

The Survey Could Have Been Simplified

One of my considerations in this research was to respect participant time and I desired to keep the survey as simple as possible. But my concern that I would not be provided access to the AEA list or would not achieve a large enough sample for the intended analysis, necessitated a number of contingency plans that complicated the survey. Questions such as whether the participant was an evaluator and an AEA member could have been removed. The question regarding language fluency was not used in the analysis; it was added after I prepared a mailing list of evaluation programs in multiple countries. Believing that I might have to distribute the survey via social media, I created a great deal of un-used survey logic complexity to permit curious non-evaluators to watch the videos without corrupting the data. The survey could have been simpler for me and for participants had I known earlier that I would have access to the AEA membership list.

Phase III: Conclusion

Phase III findings support that all three models are appropriate in evaluation, add value to the *efficiency rationale* of data viz, make sense to evaluators, and are considered useful in evaluation. Given the large overlap between the models, these complementary models do not need to be combined into an integrated model of data viz. They can be each be used in various ways and evaluators can find value in using portions of each model and combining models as they see fit to meet their needs.

Overall, findings from Phase III suggest that the *explain* ←→ *explore* model is a simple framework that an evaluator can use to consider the purpose of a particular visual before beginning to design and the *data* → *insight* model is a linear description of how to get the most information and insight out of a particular data viz. The *audience engagement* model is a holistic approach to thinking through the relationships in the evaluation to support evaluation use. While there is no clear hierarchy of models suggested in this study, comments supported that the *audience engagement* model is the most specific to evaluation – to the extent that it might not even be data viz specific.

General Discussion

This study showed that evaluators are familiar with and accept the *efficiency rationale* of data viz and that they conceptualize a number of reasons for incorporating data viz into evaluation beyond that *efficiency rationale*. This dissertation found support for incorporating a model of *data* → *insight* from the fields of computer science and cognitive science into evaluation and suggested an emerging model of data visualization specific to evaluation, the *audience engagement* model.

Evaluators Are Familiar with and Accept the Efficiency Rationale

An underlying premise of this research is that the *efficiency rationale* is a sufficient reason to use data viz but there are benefits to thinking of data viz *use* beyond simple efficiency. All three phases of this dissertation have supported that evaluators are aware of and accept the *efficiency rational* of data viz. Further, Phase III showed that participants feel that all three models presented to them “add value” to the *efficiency rationale* and suggested many uses for the models in addition to the *efficiency* rationale.

The Data → Insight Model is Appropriate and Useful in Evaluation

While the *explain* ←→ *explore* model is often reference in evaluation, the *data* → *insight* model is not commonly discussed in evaluation – it is more prominent in both computer science and cognitive science. As expected in the transdisciplinary field of evaluation (Scriven, 1994), evaluators found an evidence-based conceptual model from another field both appropriate and useful in evaluation, as it is "adaptable to many situations." Aspects of the *data* → *insight* model were already familiar to them, although they did not use the same terminology. However, I was surprised at how strongly they connected this model to evaluation activities. Participants noted that the *data* → *insight* model is both clear and linear and suggested a number of applications within evaluation.

Importantly, they also connected the model to evaluation practice overall. Comments suggested that the *data* → *insight* model is "linked directly to the purpose of program evaluation … providing insight into the how and why of a program" and that the model helps justify evaluation with stakeholders because it "makes it clear that we are not just collecting information for the sake of having data, but because we want to build upon the data to get to the point of insight, where data can be used for decision making." This conceptual framework of organizing and connecting data to create information, adding more data and information to create knowledge, and so forth, until ultimate goals are reached, appears to resonate with evaluators. There are a number of aspects of the *data* → *insight* model that I could not cover in depth in the brief video that may also resonate with evaluators. Furthermore, there are doubtless other conceptual models, not covered in this research, but well known in other fields, that could be benefit evaluation if they were introduced.

Beyond justifying evaluation, it was suggested that the *data* → *insight* model is a justification for the time needed "to use data viz to help interpret and share results." While this theme of "justifying data viz" could be extracted from other models, with effort, it was most closely associated with this model. Now that they have made this connection for me, I can see a clear connection between the evaluation specific concept of conceptual process use whereby the individual is changed based on the evaluation process and the cognitive science conceptualization of *insight*, where brain reward chemicals generate that "ah ha" moment and can physically aid in memory and internalization of knowledge.

The strongest theme in the *data* → *insight* model is that it is intuitive and easy to understand (21.6%). Taken in the context of the many comments indicating that a strength of the model is its "connection to cognitive processes" (9.8%), I believe that the model's grounding in cognitive science is related to people feeling that the model is "intuitive" and increases perceived credibility of the model. This, in turn, suggests that uncovering additional evidence to support and describe the other models in evaluation can increase their credibility.

Audience Engagement as a Conceptual Model

A model of using data viz for *audience engagement* and to extend evaluation emerged during Phase I, was articulated in Phase II, and presented to evaluators for comment in Phase III. The methodological challenges in defining the emerging *audience engagement* model well enough for inclusion in this project have already been described, but inclusion in this project is not a claim that the model is described well enough for general presentation to the field. While it is clearly aligned with evaluation practice and could be further developed into a training tool, it is not necessarily data viz specific and there are still relationships that require further definition.

I expressed concern in both Phase I and Phase II over the lack of clarity in the perceived relationship(s) between data viz, satisfaction, credibility, and professionalism. This concern was mirrored in Phase III as 10.4% of comments regarding weaknesses of the *audience engagement* model specifically mentioned concerns over the model's linkage between professionalism and credibility with data viz. Comments suggested that the "professionalism piece is over emphasized" and questioned the role of data viz in the relationship because there are "other factors at play." One even suggested the opposite relationship; data viz can hurt credibility because "in some of my evaluations, the use of data viz was seen as less scientific and credible, and the stakeholders preferred text."

Given the small number of follow-up interviews in Phase II, it is entirely possible that the professionalism component is simply over emphasized in the model or that there is a missing connection, not yet described. Further work is warranted to determine this. However, comments, such as, "it takes into account the reality of evaluation as an industry, not just a science," suggest that maybe there really is an important relationship between satisfaction and credibility in evaluation that is not palatable to some evaluators. I concur with the comment, "I wouldn't want to see an evaluator trying too hard to please an audience just to generate additional contracts," as it introduces a host of ethical concerns. Balancing the need to provide rigorous, high-quality, trustworthy findings that are *used* with running a business reliant on relationships for future contracts is a challenge for many evaluators with or without the *audience engagement* model. Setting the role of data viz aside, perhaps additional attention to this relationship is warranted.

Evaluators Use Data Viz to Explore

The cognitive efficiency and pattern recognition benefits of data viz can be harnessed at both ends of the *explain* ←→ *explore* continuum, yet a review of literature and my own experience in evaluation suggested that evaluators more often use data viz to *explain* evaluation findings. Knowing that Batch (2018) and others have suggested that sensemaking is happening during data cleaning and organization, I suggested that it is possible that evaluators are conducting exploratory data visualization during data cleaning (e.g., generating scatterplots or histograms during statistical analysis, or color-coding variables during data cleaning) and simply not conceptualizing this specifically as an *explore* data viz activity. The number of examples provided after I primed participants to conceptualize data viz as an *explore* activity for evaluators supports that evaluators are using data viz for this purpose. This, in turn, suggests that applying the label “data viz” to *explore* activities and expanding the emphasis beyond reporting provides additional opportunities to incorporate data viz into evaluation. This conceptualization of data viz to support the evaluator’s process, in addition to the audience focus already apparent in evaluation, may require additional description, training, and support within the profession.

Evaluator Comments Included Examples of Data Viz of Qualitative Data

Evaluators also use data viz for qualitative data. Phase I discussed the unexpected finding that more examples of data viz *use* were provided than expected. The study design requested reflection on “specific best work” and asked for specific examples of *use*, but it was still a surprise to receive so many concrete and specific examples of concepts often discussed in abstract terms. Eliciting specific examples of using data viz for qualitative data was not a goal of this research, but many examples were provided. Participants discussed using data viz to describe program processes, assist clients in understanding the evaluation process, and even as part of the contracting phase to explain what was going to happen next.

While the literature review did discuss visualizations of qualitative data, the examples that described program logic models and theories of change in this context were removed from the videos presented in Phase III to meet time and complexity constraints. Visualizing qualitative data is given less attention and is likely more difficult because it is further along the *data* → *insight* continuum, requiring significant human pre-processing to move it through the *information* and *knowledge* stages. Given that participants were not primed to discuss qualitative data in this way, it is interesting that so many examples emerged in this research. Evaluators clearly do consider visualizing qualitative data as “data viz” in the context of this study but were given limited opportunity to discuss it as such. Expanding conceptualizations, descriptions, training, and support on visualizing qualitative data within the profession could be of great benefit considering the large amount of qualitative data used in evaluation.

No Model to Rule Them All

Early in this research I anticipated that a future direction would be to conduct further research with the goal of creating a combined model of data viz *use*. At this time, I do not think it is practical nor valuable to do so. Evaluation is steeped in context and often requires unique and targeted approaches. Evaluators in this research found value in each and every one of the models. The comments demonstrated that professional evaluators are adept at taking what they need from any particular model and applying it to fit the situation. I do not see a need for a comprehensive and prescriptive model at this stage. I think there is tremendous value in describing these, and potentially other, evidence-based conceptual frameworks to evaluators and allowing them to apply portions as needed.

It is true that not every participant appreciated the abstraction of the videos. The model is "just a way of thinking and does not say 'how' to use it" was a weakness identified for both the *explain* ←→ *explore* model (n = 8) and the *data* → *insight* model (n = 3). This is consistent with the opposing comments that the same model is either "too simple" or not simple enough, "too detailed" or not detailed enough, etc. Recognizing that no model will be perfect for all evaluators nor all situations, the strength of these models is that they overlap and can be applied in part or in whole as needed for each circumstance.

This Study Provided Knowledge in Exchange for Participant Time

Out of respect for the evaluation professionals volunteering their time on this survey, one goal of this study design was to give survey participants knowledge in exchange for their time. More than 75% of participants responded *agree* or *strongly agree* that they "learned something new" about the first model they watched. Particularly in the case of the *data* → *insight* and *audience engagement* models, participants indicated they had "learned something new" even though they were already familiar with the models. Even after watching multiple model videos, they continued agreeing that they had learned something new, suggesting that they were continuing to learn from subsequent videos, despite overlap between the models. This, combined with the number of participants that continued the survey to watch two videos (n = 62; 47.3%) or three videos (n = 25; 19.1%), suggests that participants found value in the models presented and were interested in learning more about data viz in this explainer video medium.

General Strengths and Contribution

The most obvious strength of this research project is that the topic – data visualization – is of interest within program evaluation and findings suggest that considering different conceptual models of data visualization is appropriate and valuable in evaluation. The conceptual models presented in this project provide program evaluators with evidence-based ways of thinking about using data viz in their work that they may not have considered before. The efficiency rationale for data viz and all three conceptual models presented in Phase III have been articulated as brief explainer videos and are now available on YouTube for any evaluator to consider. Through this dissemination, evaluators will be offered ways of thinking about data viz use that they may not have previously considered and can now combine in new ideas.

Ultimately, from these various models, evaluators may glean new ideas for communicating and building data viz capacity within their clients. Clearly present within the *audience engagement* model is an emphasis on audience satisfaction which includes building audience or client capacity and giving the client data viz that the client can then use with additional audiences. Similarly, a better understanding of the varied reasons – beyond *efficiency* and *explain* – for using data viz can be disseminated from evaluators to their clients such that they can take advantage of the multiple and context specific reasons for using data viz. This has the potential to extend evaluation use to distal audiences.

Academically, the primary strength of this research project lies in the mixed methods design. Qualitative methods provided a better understanding of the conceptualizations of data viz use in program evaluation and elucidated the emerging model of *audience engagement*. Quantitative methods were used to measure awareness and perceived usefulness of these models, and then qualitative methods were used to enhance the findings from those quantitative findings.

A less obvious strength of this project, but of particular interest to me, is that it was designed to emphasize respect for persons in the research process by using existing data when possible and by creating teaching “artifacts” of the project that can be shared with less research-focused evaluation audiences. Having participated in many research projects that were under disseminated, I endeavored to produce at least some small contribution that could be used easily without formal publication processes.

General Conclusion

Many resources, including many program evaluation specific resources, explain *why* and *how-to* create data viz (e.g., Evergreen, 2017, 2020; Few, 2012, 2019; Kirk, 2019). The majority of these sources rely on the *efficiency rationale* and this rationale is likely sufficient most of the time. The goal of this dissertation was to consider the purposes that evaluators can achieve with that efficiency. This research supported that program evaluators accept the *efficiency rationale* and are interested in other reasons for using data viz beyond *efficiency*.

Program evaluators in this sample do understand and accept the *efficiency rationale* of data viz. They *usually* (38.2%) or *always* (42.7%) use data viz in their evaluation work. At least once per week, 90% of participants *see* a chart, over half *use* data viz to make a decision, and over half *create* data viz. Learning new ways of displaying data is less common, with the majority of participants (87.1%) doing so once a month or less. However, participation in this voluntary project and the large number of participants who watched multiple videos suggests that there is desire to learn more about data viz in evaluation. Furthermore, the interest expressed in the conceptual models and their deftness at applying portions of the models to fit their needs, suggests that evaluators are interested in and find value in learning experiences that cover abstract concepts in data viz. Interest in learning about data viz is a subtle but important distinction from the technical or "how to create" data viz content most frequently offered. The interest in both the content and the medium (e.g., brief explainer videos) suggests new opportunities for teaching experiences and professional development opportunities within the profession.

Interviews showed that evaluators conceptualize and discuss reasons for using data viz consistent with a *Utilization-Focused Framework*, the *explain* ←→ *explore* model, the *data* → *insight* model, and an emerging *audience engagement* model. Upon presenting three of these conceptual models (*explain* ←→ *explore*, *data* → *insight*, and *audience engagement)* to a sample of evaluators who are AEA members, it was determined that all three frameworks are complementary, appropriate, and useful in evaluation, and have the potential to benefit evaluators as they consider why they should use data visualization in their work. Importantly, data viz is a time-consuming skill that may not be of interest to all evaluators, but providing evaluators with conceptualizations of data viz beyond *efficiency* may make them more willing to expend the time and effort needed to apply data viz to their evaluation work. While only one tool among many, and not appropriate for all situations, data viz can provide efficiency gains over text-based communication while offering the other benefits described in the models.

Future Directions

As all models were deemed appropriate and useful in evaluation, the next step of this research is to disseminate the models more widely among evaluators. Continuing to elicit feedback on the models and improve them can serve both evaluation theory and practice.

Findings across all three phases suggest that there is interest, need, and desire for more professional development in data visualization and associated skills. Beyond skills workshops and "how to" guides, findings also suggest a desire for more learning opportunities about abstract concepts. Conferences, workshops, textbooks, and online trainings are available to evaluators, but the majority of this sample reports learning new ways to create viz one time per month or less. It would be valuable to consider whether offering smaller trainings in other media can play a role in ongoing professional development.

The *U-FE* model was removed from this study as it could not be simplified enough to meet the self-imposed requirement that videos be less than four-minutes, but it describes principles and a number of conceptualizations that are likely valuable to evaluators. Redesigning and releasing key components of that model, then collecting feedback to improve and extend it is a future direction.

The most obvious follow-up project after this dissertation is to continue defining the *audience engagement* model. The existing model needs to be disseminated more broadly for feedback from evaluators and efforts need to be made to better understand a number of relationships associated with the model. For instance, the presumed connection between credibility and professionalism is ill-defined in the model but this relationship likely exists – and likely exists separate from data viz such that the research needed to explicate it may not even involve data viz.

There was also a strong emphasis in the *audience engagement* model on evaluation as a profession and efforts to elevate evaluation overall. Additional research into understanding how this concept overlaps with efforts to define and professionalize evaluation is warranted. There were several other themes that were beginning to emerge through these discussions that readily lend themselves to more questions. For instance, these interviewees appeared to tolerate high levels of ambiguity over their definitions of *use* and accepted the nebulous nature of their descriptions. Relatedly, they exuded a sense of optimism that their efforts will not be in vain, that their evaluation contributions will eventually be used, and that even the most distal *use* is a form of *use*. Additional research on these topics could benefit evaluation whether considered through the lens of data viz or not.

References

Abbato, S. (2023). Digital Evaluation Stories: A Case Study of Implementation for Monitoring and Evaluation in an Australian Community Not-for-Profit. *American Journal of Evaluation, 44*(4), 604-628. https://doi.org/10.1177/10982140221138031

Alkin, M. C., & King, J. A. (2016). The historical development of evaluation use. *American Journal of Evaluation, 37*(4), 568-579. https://doi.org/10.1177/1098214016665164

Alkin, M. C., & King, J. A. (2017). Definitions of evaluation use and misuse, evaluation influence, and factors affecting use. *American Journal of Evaluation, 38*,(3), 434-450. https://doi.org/10.1177/1098214017717015

Alkin M. C., Taut S. M. (2002). Unbundling evaluation use. *Studies in Educational Evaluation, 29*(1), 1-12. https://doi.org/10.1016/S0191-491X(03)90001-0

American Evaluation Association. (2018). *Evaluator competencies.* https://www.eval.org/About/Competencies-Standards

Aparicio, M., & Costa, C. J. (2014). Data visualization. *Communication Design Quarterly, 3*(1), 7-11. https://doi.org/10.1145/2721882.2721883

Assarroudi, A., Heshmati Nabavi, F., Armat MR, Ebadi A., & Vaismoradi, M. (2018). Directed qualitative content analysis: The description and elaboration of its underpinning methods and data analysis process. *Journal of Research in Nursing, 23*(1), 42-55. https://doi.org/10.1177/1744987117741667

Azzam, T., Evergreen, S., Germuth, A. A., & Kistler, S. J. (2013). Data visualization and evaluation. *New Directions for Evaluation, 2013*(139), 7-32. https://doi.org/10.1002/ev.20065

Azzam, T., Douville, S., Knight, C., Grandjean Targos, P., & Jones, N.D. (in press). Information visualization and evaluation. In *The handbook of program evaluation*.

Bamberger, M., & Mabry, L. (2020). *Realworld evaluation: Working under budget, time, data, and political constraints.* SAGE Publications.

Batch, A., & Elmqvist, N. (2018). The interactive visualization gap in initial exploratory data analysis. *IEEE Transactions on Visualization and Computer Graphics, 24*(1), 278-287. https://doi.org/10.1109/TVCG.2017.2743990

Bertin, J. (1977). *Graphics and graphic information processing*. de Gruyter Press.

Börner, K., Bueckle, A., & Ginda, M. (2019). Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proceedings of the National Academy of Sciences of the United States of America, 116*(6), 1857-1864. https://doi.org/10.1073/pnas.1807180116

Braskamp, L. A. (1982). A definition of use. *Studies in Educational Evaluation, 8*, 169-174. https://doi.org/10.1016/0191-491X(82)90009-8

Brehmer, M., & Munzner, T. (2013). A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics, 19*(12), 2376-2385. https://doi.org/10.1109/TVCG.2013.124

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: Using vision to think.* Morgan Kaufmann Publishers.

Cavaller, V. (2021). Dimensional taxonomy of data visualization: A proposal from communication sciences tackling complexity. *Frontiers in Research Metrics and Analytics, 6*, 1-22. https://doi.org/10.3389/frma.2021.643533

Centers for Disease Control and Prevention. (1999). *Framework for program evaluation in public health.* https://www.cdc.gov/evaluation/framework/index.htm

Chang, R., Ziemkiewicz, C., Green, T. M., & Ribarsky, W. (2009). Defining insight for visual analytics. *IEEE Computer Graphics and Applications*, 14-17. https://doi.org/10.1109/mcg.2009.22

Chen, M., Ebert, D., Hagen, H., Laramee, R. S., van Liere, R., Ma, K., Ribarsky, W., Scheuermann, G., & Silver, D. (2009). Data, information, and knowledge in visualization. *IEEE Computer Graphics and Applications, 29*(1), 12-19. http://dx.doi.org/10.1109/MCG.2009.6

Chen, H. T. (2015). *Practical Program Evaluation: Theory-Driven evaluation and the integrated evaluation perspective* (2nd ed.). SAGE Publications.

Cheng, H. G., & Phillips, M. R. (2014). Secondary analysis of existing data: Opportunities and implementation. *Shanghai Archives of Psychiatry, 26*(6), 371-375. https://doi.org/10.11919/j.issn.1002-0829.214171

Clayton, M. (2014). The Origin of Stakeholders. In: *The Influence Agenda*. Palgrave Macmillan.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Psychology Press.

Conati, C., Carenini, G., Hoque, E., Steichen, B., & Toker, D. (2014). Evaluating the impact of user characteristics and different layouts on an interactive visualization for decision making. *2014 Eurographics Conference on Visualization (EuroVis), 33*(3). https://doi.org/10.1111/cgf.12393

Corbin, J., & Strauss, A. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (4th ed.). SAGE Publications.

Cousins, J.B. (2007). Editor's notes. *New Directions for Evaluation*, 1-4. https://doi.org/10.1002/ev.239

Creswell, J. (2009). *Research design: Qualitative, quantitative and mixed methods approaches* (3rd ed.) SAGE Publications.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry & research design* (4th ed.). SAGE Publications.

Dasu, T., & Johnson, T. (2003). *Exploratory data mining and data cleaning.* John Wiley & Sons, Inc.

Dean-Coffey, J. (2013). Graphic recording. *New Directions for Evaluation, 2013*(140), 47–67. https://doi.org/10.1002/ev.20073

Donaldson, S. I. (2007). *Program theory-driven evaluation science: Strategies and applications.* Taylor & Francis Group, LLC.

Douville, S. (2019). *Prevalence and early warning signs of evaluation misuse* [Unpublished master's thesis]. Claremont Graduate University.

Douville, S., Grandjean Targos, P., Jones, N. D., Knight, C., & Azzam, T. (Submitted). Data visualization expert lessons learned: Implications for program evaluators. *American Journal of Evaluation*.

Evergreen, S. D. H. (2011). *Death by boredom: The role of visual processing theory in written evaluation communication*. [Doctoral dissertation, Western Michigan University.] https://scholarworks.wmich.edu/dissertations/403

Evergreen, S. D. H. (2017). *Effective data visualization: The right chart for the right data.* SAGE Publications.

Evergreen, S. D. H. (2018). *Presenting data effectively: Communicating your findings for maximum impact* (2nd ed.). SAGE Publications.

Evergreen, S., & Metzner, C. (2013). Design principles for data visualization in evaluation. *New Directions for Evaluation, 2013*(140), 5-20. https://doi.org/10.1002/ev.20071

Evergreen, S., & Sabarre, N. (2019). Branding for the independent consultant: Basic to advanced. *New Directions for Evaluation, 164*, 101-113. https://doi.org/10.1002/ev.20392

Few, S. (2012). *Show me the numbers: Designing tables and graphs to enlighten*. Analytics Press.

Few, S. (2019). *The data loom: Weaving understanding by thinking critically and scientifically with data.* Analytics Press.

Freedman, E. G., & Shah, P. (2002). Toward a model of knowledge-based graph comprehension. In M. Hegarty, B. Meyer, & N. H. Narayanan (Eds.), *Diagrams 2002: Diagrammatic Representation and Inference - Second International Conference, April 18-20, 2002, Callaway Gardens, Georgia, USA* (pp. 18-30). https://link.springer.de/link/service/series/0558/bibs/2317/23170018.htm

Glaser, B. G. (1965). The constant comparative method of qualitative analysis. Social Problems, 12(4), 436-445. https://psycnet.apa.org/doi/10.1525/sp.1965.12.4.03a00070

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Sociology Press.

Hegarty, M. (2011). The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science, 3*(3), 446-474. https://doi.org/10.1111/j.1756-8765.2011.01150.x

Henderson, S., & Segal, E. H. (2013). Visualizing qualitative data in evaluation research. *New Directions for Evaluation, 139*, 53–71. https://doi.org/10.1002/ev.20067

Hennink, M., & Kaiser, B. N. (2022). Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science and Medicine, 292*(114523). https://doi.org/10.1016/j.socscimed.2021.114523

Hennink, M. M., Kaiser, B. N., & Marconi, V. C. (2017). Code saturation versus meaning saturation: How many interviews are enough?. *Qualitative Health Research, 27*(4), 591–608. https://doi.org/10.1177/1049732316665344

Holder, E. & Xiong, C. (2022). Dispersion vs disparity: Hiding variability can encourage stereotyping when visualizing social outcomes. *IEEE Transactions on Visualization and Computer Graphics, 29*(1), 624-634. https://doi.org/10.1109/TVCG.2022.3209377

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277-1288. https://doi.org/10.1177/1049732305276687

Huang, W., Eades, P., & Hong, S. H. (2009). Measuring effectiveness of graph visualizations: A cognitive load perspective. *Information Visualization, 8*(3), 139-152. https://doi.org/10.1057/ivs.2009.10

Isenberg, T., Isenberg, P., Chen, J., Sedlmair, M., & Moller, T. (2013). A systematic review on the practice of evaluating visualization. *IEEE Transactions on Visualization and Computer Graphics, 19*(12), 2818-2827. https://doi.org/10.1109/TVCG.2013.126

Johnson, K., Greenseid, L. O., Toal, S. A., King, J. A., Lawrenz, F., & Volkov, B. (2009). Research on evaluation use: A review of the empirical literature from 1986 to 2005. *American Journal of Evaluation, 30*(3), 377-410. https://doi.org/10.1177/1098214009341660

Jones, N. D., Azzam, T., Wanzer, D. L., Skousen, D., Knight, C., & Sabarre, N. (2019). Enhancing the effectiveness of logic models. *American Journal of Evaluation, 40*(1), 1-19. https://doi.org/10.1177/1098214018824417

Kandogan, E., Balakrishnan, A., Haber, E. M., & Pierce, J. S. (2014). From data to insight: Work practices of analysts in the enterprise. *IEEE Computer Graphics and Applications, 34*(5), 42-50. https://doi.org/10.1109/MCG.2014.62

Kerr, C., Nixon, A., & Wild, D. (2010). Assessing and demonstrating data saturation in qualitative inquiry supporting patient-reported outcomes research. *Expert Review of Pharmacoeconomics & Outcomes Research, 10*(3), 269–281. https://doi.org/10.1586/erp.10.30

Kirk, A. (2019). *Data visualisation: A handbook for data driven design* (2nd ed.). SAGE Publications.

King, J. A., & Alkin, M. C. (2019). The centrality of use: Theories of evaluation use and influence and thoughts on the first 50 years of use research. *American Journal of Evaluation, 40*(3) 431-458. DOI: 10.1177/1098214018796328

Kirkhart, K. E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation, 88*(Winter), 5-23. https://doi.org/10.1002/ev.1188

Klein, G., Moon, B., & Hoffman, R. R. (2006). Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent Systems*, *21*(5). https://doi.org/10.1109/MIS.2006.100

Knorr, K. D. (1977). Policymakers' use of social science knowledge: Symbolic or instrumental? In C. Weiss (Ed.), *Using social research in public policy making*, 165-182. D. C. Health.

Kosara, R. (2007). Visualization criticism – The missing link between information visualization and art. *11th International Conference Information Visualization (IV '07)*, 631-636. https://doi.org/10.1109/IV.2007.130

Kujala, J., Sachs, S., Leinonen, H., Heikkinen, A., & Laude, D. (2022). Stakeholder engagement: Past, present, and future. *Business & Society, 61*(5), 1136-1196. https://doi.org/10.1177/00076503211066595

Laerd Statistics. (2024). *Chi-square goodness-of-fit test in SPSS statistics.* https://statistics.laerd.com/spss-tutorials/chi-square-goodness-of-fit-test-in-spss-statistics.php

Laerd Statistics. (2024). *Chi-square test for association using SPSS statistics.* https://statistics.laerd.com/spss-tutorials/chi-square-test-for-association-using-spss-statistics.php

Laerd Statistics. (2024). *One-way ANOVA in SPSS statistics.* https://statistics.laerd.com/spss-tutorials/one-way-anova-using-spss-statistics.php

Lee, S., Kim, S. H., Hung, Y. H., Lam, H., Kang, Y., & Yi, J. S. (2014). How do people make sense of unfamiliar visualizations?: A grounded model of novice's information visualization sensemaking. *IEEE Transactions on Visualization and Computer Graphics, 22*(1), 499-508. https://doi.org/10.1109/tvcg.2015.2467195

Li, Y., & Xie, Y. (2020). Is a picture worth a thousand words? An empirical study of image content and social media engagement. *Journal of Marketing Research, 57*(1), 1-19. https://doi.org/10.1177/0022243719881113

Long, A., Jennings, J., Bademosi, K., Chandran, A., Sawyer, S., Schumacher, C., Greenbaum, A., & Fields, E. L. (2022). Storytelling to improve healthcare worker understanding, beliefs, and practices related to LGBTQ patients: A program evaluation. *Evaluation and Program Planning, 90*. https://doi.org/10.1016/j.evalprogplan.2021.101979

Lysy, C. (2013). Developments in quantitative data display and their implications for evaluation. *New Directions for Evaluation, 139*, 33–51. https://doi.org/10.1002/ev.20066

MacDonald, G., & McLees, A. (2021, August 3). As an evaluator, do I use words (e.g., stakeholder) that can be harmful to others? *AEA365.org blog post*. Retrieved 12/19/2023 from https://aea365.org/blog/as-an-evaluator-do-i-use-words-e-g-stakeholder-that-can-be-harmful-to-others-by-goldie-macdonald-anita-mclees

Mandinach, E. B., Honey, M., & Light, D. (2006). A theoretical framework for data-driven decision making. *Annual Meeting of the American Educational Research Association, San Francisco, CA (April)*.

Mark, M. M., & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation, 10*(1), 35-57. https://doi.org/10.1177/1356389004042326

Mason, S., & Azzam, T. (2019). In need of an attitude adjustment? The role of data visualization in attitude change and evaluation influence. *American Journal of Evaluation, 40*(2), 249-267. https://doi.org/10.1177/1098214018778808

McAlindon, K., Watling Neal, J., Neal, Z. P., Mills, K. J., & Lawlor, J. (2019). The BOND framework: A practical application of visual communication design and marketing to advance evaluation reporting. *American Journal of Evaluation, 40*(2), 291-305. https://doi.org/10.1177/1098214018771219

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). SAGE Publications.

Moore, J. (2017). Data visualization in support of executive decision making. *Interdisciplinary Journal of Information, Knowledge, and Management, 12*, 125-138. http://www.informingscience.org/Publications/3687

North, C. (2006). Toward measuring visualization insight. *IEEE Computer Graphics and Applications, 26*(3), 6-9. https://doi.org/10.1109/MCG.2006.70

Nussbaumer Knaflic, C. N. (2015). *Storytelling with data: A data visualization guide for business professionals.* John Wiley & Sons, Inc.

Oxford Dictionary. (n.d.) The Oxford English Dictionary Online. Retrieved October 15, 2022 and December 20, 2023.

Pankaj, V., & Emery, A. K. (2016). Data placemats: A facilitative technique designed to enhance stakeholder understanding of data. *New Directions for Evaluation, 149*, 81–93. http://dx.doi.org/10.1002/ev.20181

Patton, M.Q. (1990). Qualitative evaluation and research methods (2nd ed.). SAGE Publications.

Patton, M. Q. (2008). *Utilization-focused evaluation.* SAGE Publications.

Patton, M. Q. (2015). *Qualitative research and evaluation methods*. SAGE Publications.

Patton, M. Q., & Campbell-Patton, C. E. (2022). *Utilization focused evaluation* (5th ed.). SAGE Publications.

Pelz, D. C. (1978). Some expanded perspectives on use of social science in public policy. In J. M. Yinger & S. J. Cutler (Eds.), *Major social issues: A multidisciplinary view* (pp. 346-357). Free Press.

Pfitzner, D., Hobbs, V., & Powers, D. (2001). A unified taxonomic framework for information visualization. In J. Weckert (Ed.), *Proceedings of the 2nd Australian Institute of Computer Ethics Conference (AICE2000)*, Canberra. *Conferences in Research and Practice in Information Technology, 1*.

Pienta, A. M., McFarland O'Rourke, J., & Franks, M. M. (2011). Getting started: Working with secondary data. In K. H. Trzesniewski, M. B. Donnellan, & R. E. Lucas (Eds.), *Secondary Data Analysis: An Introduction for Psychologists*. American Psychological Association. http://dx.doi.org/10.1037/12350-001

Pieters, R., & Wedel, M. (2004). Attention capture and transfer in advertising: Brand, pictorial, and text-size effects. *Journal of Marketing, 68*(2), 36-50. https://doi.org/10.1509/jmkg.68.2.36.27794

Pirolli, P., & Card, S. K. (2005). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of the International Conference on Intelligence Analysis*.

Plano Clark, V. L., & Sanders, K. (2015). The use of visual displays in mixed methods research: Strategies for effectively integrating the quantitative and qualitative components of a study. In M. T. McCrudden (Ed.), *The Use of Visual Displays in Research and Testing: Coding, Interpreting, and Reporting Data* (pp. 177-206).

Rich, R.F. (1977). Measuring knowledge utilization: Processes and outcomes. *Knowledge and Policy, 10*, 11–24. https://doi.org/10.1007/BF02912504

Roth, R. E. (2012). An empirically-derived taxonomy of cartographic interaction primitives. *Proceedings of GIScience*, 18-21.

Rykaczewska, A. M. H. (2021). *Achieving evaluation influence through elaboration likelihood model-informed evaluation product designs* [Doctoral dissertation, Claremont Graduate University].

Sandelowski, M., Voils, C. I., & Knafl, G. (2009). On quantitizing. *Journal of Mixed Methods Research, 3*(3), 208-222. https://doi.org/10.1177/1558689809334210

Scriven, M. (1994). Evaluation as a discipline. *Studies in Educational Evaluation, 20*, 147-166.

Shah, P., & Carpenter, P. (1995). Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology General, 124*, 337-370. http://dx.doi.org/10.1037/0096-3445.124.1.43

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. *Proceedings 1996 IEEE Symposium on Visual Languages*, 336-343. https://doi.org/10.1109/VL.1996.545307

Slone, D. J. (2009). Visualizing qualitative information. *The Qualitative Report, 14*(3), 488-497. https://doi.org/10.46743/2160-3715/2009.1407

Smith, V. S. (2013). Data dashboard as evaluation and research communication tool. *New Directions for Evaluation, 140*, 21-45. https://doi.org/10.1002/ev.20072

Stufflebeam, D. L. (1980). An interview with Daniel L. Stufflebeam. *Educational Evaluation and Policy Analysis, 2*(4) 90-92. https://doi.org/10.3102%2F01623737002004085

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*, 261-292. https://doi.org/10.1007/s10648-019-09465-5

Taylor, S. J., Bogdan, R., & DeVault, M. (2016). *Introduction to qualitative research methods: A guidebook and resource* (4th ed.). Wiley & Sons Inc.

Tik, M., Sladky, R., Luft, C. D. B., Willinger, D., Hoffmann, A., Banissy, M. J., Bhattacharya, J., & Windischberger, C. (2018). Ultra-high-field fMRI insights on insight: Neural correlates of the Aha!-moment. *Human brain mapping, 39*(8), 3241-3252. https://doi.org/10.1002/hbm.24073

Tufte, E. R., (2001). *The visual display of quantitative information* (2nd ed.). Graphics Press.

Tufte, E. R. (2016, September). Keynote address to the Microsoft Machine Learning and Data Science Summit. Retrieved from https://www.youtube.com/watch?v=rHUDJ8RyseQ

Valéry, R. (2007). Are program evaluators judges and/or knowledge brokers? *Journal of Epidemiology and Community Health, 61*(12). https://doi.org/10.1136/jech.2007.065573

Verdinelli, S., & Scagnoli, N. I. (2013). Data display in qualitative research. *International Journal of Qualitative Methods, 12*, 359-381. https://doi.org/10.1177/160940691301200117

Wanzer, D. L., Azzam, T., Jones, N. D., & Skousen, D. (2020). The role of titles in enhancing data visualization. *Evaluation and Program Planning, 84.*

Ware, C. (2012). *Information visualization: Perception for design* (3rd ed.). Elsevier.

Ware, C. (2021). *Information visualization: Perception for design* (4th ed.). Elsevier.

Weiss, C. H. (Ed.). (1977). *Using social research in public policy making.* Lexington Books.

Weiss, C. H. (1998). *Evaluation* (2nd ed.). Prentice-Hall, Inc.

Weissgerber, T. L., Savic, M., Winham, S. J., Stanisavljevic, D., Garovic, V. D., Milic, N. M. (2017). Data visualization, bar naked: A free tool for creating interactive graphics. *Journal of Biological Chemistry, 292*(50), 20592-20598. https://doi.org/10.1074/jbc.RA117.000147

W. K. Kellogg Foundation. (2004). *Using logic models to bring together planning, evaluation, and action: Logic model development guide.* Retrieved from: https://wkkf.issuelab.org/resource/logic-model-development-guide.html

Yi, J. S., Kang, Y., Stasko, J. T., & Jacko, J. A. (2008). Understanding and characterizing insights: How do people gain insights using information visualization? *BELIV '08: Proceedings of the 2008 Workshop on BEyond time and errors: Novel evaLuation methods for Information Visualization,* 4, 1-6. https://doi.org/10.1145/1377966.1377971

Zhu, Y. (2007). Measuring effective data visualization. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, N. Paragios, S. M. Tanveer, T. Ju, Z. Liu, S. Coquillart, C. Cruz-Neira, T. Müller, & T. Malzbender (Eds.), *Advances in visual computing: ISVC 2007 Lecture Notes in Computer Science* (Vol. 4842, pp. 652-661). Springer-Verlag. https://doi.org/10.1007/978-3-540-76856-2_64

Appendix A: 2021 Interview Team and Video Review Team Bios

Sarah Douville

ORCID: 0000-0002-0134-058X

Sarah Douville has a business management and finance background with fifteen years of experience in program evaluation and data visualization of academic healthcare program data. As she currently owns and manages a small software as a service company which includes data visualization in its product offerings, she is primed to focused on software, technology, automation, and scalability within data visualization. Her recent interests are concerned with audience visualcy, trade-offs between aesthetics and efficiency in visualization, and techniques for visualizing qualitative data. Sarah is completing a PhD focused on program evaluation at Claremont Graduate University.

Piper T. Grandjean Targos

ORCID: 0000-0002-5335-7762

Piper T. Grandjean Targos owns and manages a private evaluation consulting company conducting a wide variety of small to mid-sized evaluation projects for private and non-profit organizations. With close client relationships and a focus on participatory capacity building work, Piper is primed to recognize concerns related to clients' perspectives and is attuned to balancing stakeholder needs within tight budgets. Piper's data visualization research combines her background in psychology and grounding in social science principles with her interests in visual arts and emphasis on visual appeal and evaluation use. Piper is completing a PhD focused on program evaluation at Claremont Graduate University.

Natalie D. Jones

ORCID: 0000-0002-5264-206X

Natalie Jones is an evaluation project manager and consultant for a large public university's evaluation department. Although her current work consists of primarily higher education academic clients, her work has spanned non-profit community based and corporate for-profit organizations. With a focus on evaluation communication and data visualization, Natalie is primed to focus on the client engagement process and perspective along with a focus on the ultimate alignment to current industry data visualization principles. Natalie is completing a PhD focused on program evaluation at Claremont Graduate University.

Ciara Knight

Ciara Knight (pronouns: she, her) is a Black educator, evaluation consultant, and researcher with expertise in culturally responsive and equitable approaches. She has over 14 years of experience as an evaluator with diverse communities and programs. Accordingly, she focuses on how to ensure data visualization is useful and credible based upon the cultural lens of the audience. Ciara has completed a PhD focused on program evaluation at Claremont Graduate University.

Tarek Azzam

ORCID: 0000-0003-3864-0217

Dr. Tarek Azzam is an Associate Professor in the Department of Education at University of California, Santa Barbara and directs the university's Evaluation Center. His research and evaluation experience includes projects related to student academic and retention programs, Science Technology Engineering Math (STEM) education programs, children's health programs, and international schools and development programs. His research is focused on integrating new technologies and resources, such as crowdsourcing, to develop new evaluation specific methodologies. With an emphasis on real-world evaluations, his work aims to improve the rigor and credibility of evaluations and increase potential impact on programs and policies.

Appendix B: 2021 Consent and Survey

Claremont Graduate University

AGREEMENT TO PARTICIPATE IN A STUDY TO BETTER UNDERSTAND THE PROCESS OF DATA VISUALIZATION (IRB #3882)

You are invited to participate in a research study consisting of one survey and one interview. Volunteering will probably not benefit you directly, but you will be helping us explore the process of creating data visualization. If you volunteer, you will be asked to (a) complete a 15-20 minute survey about your experience with data visualization, and (b) participate in a 45-60 minute interview discussing the process you have used when creating data visualization. The entire study will take about 60-90 minutes of your time. Volunteering for this study involves no more risk than what a typical person experiences on a regular day. Your involvement is entirely up to you. You may withdraw at any time for any reason. Please continue reading for more information about the study.

Study Leadership: This research study is led by Tarek Azzam, a professor at Claremont Graduate University with aid from the data visualization lab.

Purpose: The purpose of this study is exploratory in nature. It seeks to build a foundational understanding of the process of creating a data visualization.

Eligibility: To be in this study you must create data visualization professionally and be able to discuss the process as you have experienced it.

Participation: During the study, you will be asked to complete one survey and participate in one interview conducted by a member of the data visualization lab. Data from this study will be analyzed and reported by the researchers in academic journal(s) in an attempt to further the field of data visualization. Participation will take between 60-90 minutes and questions will ask about your experience and process for creating data visualization. Some questions you will be asked are, for example, tell me briefly about a data visualization you have created? What was the project? How did you begin this process?

Risks Of Participation: The risks you run by taking part in this study are minimal and are not greater than those ordinarily encountered in daily life or during the performance of routine work.

Benefits Of Participation: We **do not** expect the study to benefit you personally. This study is intended to benefit the data visualization community by informing them as to common practices in data visualization.

Compensation: You will not be directly compensated for participating in this study. Your participation is greatly appreciated.

Voluntary Participation: Your participation in this study is completely voluntary. You may stop or withdraw from the study at any time without it being held against you. You may also refuse to answer any question at any time. Your decision whether or not to participate will have no effect on your current or future connection with anyone at CGU.

Confidentiality: Your individual privacy will be protected in all papers, books, talks, posts, or stories resulting from this study. We may share the data we collect with other researchers, but we will not reveal your identity with it. In order to protect the confidentiality of your responses, we will assign each participant a pseudonym. Files containing your true identity, such as your consent form and audio/video recordings will be stored in password protected accounts accessed only by the research team.

At the beginning of the interview we will ask to digitally record the conversation. These digital recordings will be saved from the recording device to a password protected shared drive immediately after the interview is completed. After checking to ensure the transition was successful, the recording will be deleted from the digital recording device.

The digital recordings will be used to produce transcripts/detailed notes for analysis. If at any point your name is mentioned during the interview, it will simply be replaced with "the interviewee" or your pseudonym when transcribed. Audio/video recordings will be destroyed after transcripts have been created and verified. Transcripts will be identified only by pseudonym and will be stored in password protected accounts accessed only by the research team.

Survey data and transcripts/detailed notes will be retained by the researchers until publications from the research study have been developed and presented in the peer-reviewed literature.

Further Information: If you have any questions or would like additional information about this study, please contact Tarek Azzam at tarek.azzam@cgu.edu, Devin Larsen at devin.larsen@cgu.edu, or Sarah Douville at sarah.douville@cgu.edu. The CGU Institutional Review Board has approved this project. If you have any ethical concerns about this project or about your rights as a human subject in research, you may contact the CGU IRB at (909) 607-9406 or at irb@cgu.edu. You may print and keep a copy of this consent form. **OR** If you wish, we will be happy to send you a copy of this consent form.

Consent: Typing your name below indicates that you understand the information on this form, that someone has answered any and all questions you may have about this study, and you voluntarily agree to participate in it. If you do not consent, simply leave the rest of the survey blank. Thanks!

What is your definition of data visualization?

What is your purpose for using data visualization?

How were you trained in data visualization?

What resources do you use to improve your data visualization skills? *Examples may include specific conferences you attend, books you reference, websites you frequent or people you follow.*

In the last year, how often has your work involved data visualization?

- ◯ Always
- ◯ Often
- ◯ Sometimes
- ◯ Rarely
- ◯ Never

How often, in the last year, was your data visualization process collaborative?

- ◯ Always
- ◯ Often
- ◯ Sometimes
- ◯ Rarely
- ◯ Never

How many years of experience do you have in data visualization?

How long has it been since you created a data visualization?

- ◯ Currently working on a visualization
- ◯ 1-6 months ago
- ◯ 7-11 months ago
- ◯ 1 year ago
- ◯ More than a year ago

What sector do you work in? *Examples may include education, business, journalism or public health.*

Who do you usually create data visualizations for?

What is your age?

What is your gender?

- Male
- Female
- Other
- Decline to state

Do you consider yourself to be:

- Asian or Pacific Islander
- Black or African American
- Hispanic or Latino
- Native American or American Indian
- White or Caucasian
- I identify as:
- Decline to state

What is the highest level of education you have completed?

- Did not complete high school
- High School/ GED

- ◯ Some college
- ◯ Bachelor's degree
- ◯ Master's degree
- ◯ Advanced graduate work or PhD

What is your academic discipline?

As this study is related to creating data visualizations, we would like to know:
Do you have any impairments (visual, visual processing, etc.?)

- ☐ No visual impairments
- ☐ Color blindness
- ☐ Low vision
- ☐ Dyslexia
- ☐ Other impairments:

Appendix C: 2021 Interview Questions

<table>
<tr><td colspan="3">I’m talking with you today because you are an expert in data visualization and I want to know more about your process for developing data visualizations. I’m going to ask about your experience with data visualization and then walk you through examples of your own data visualization experiences. So, first,</td></tr>
<tr><td>5 min</td><td>1</td><td>Tell me a little about your path to data visualization. What originally prompted your interest in data visualization?</td></tr>
<tr><td colspan="3">Thank you! Now, I want to walk through a specific example of a data visualization you created and dig into that experience. You have provided a data visualization that you believe represents your best work.

[*Thank them for sending it ahead of time or ask to review it on screen. If that is not possible, ask for a clear description of the viz.*]</td></tr>
<tr><td rowspan="3">5 min</td><td>2</td><td>First, tell me briefly about this specific visualization project. What was the project and what did you create?

[*Please describe the final viz for me. Have them narrate even if you can see the viz so we have a verbal record. As in: This viz won’t be available to others, so could you please explain in words what we are looking at right now?*]</td></tr>
<tr><td>3</td><td>What was the purpose and audience for the viz?</td></tr>
<tr><td>4</td><td>Why do you believe that this represents your best work?</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td rowspan="2">25 min</td><td>5</td><td>I'm going to ask about your process in a minute, but first I want to know everything about the context of that project.

[*Summarize for them the situational factors you already heard and then ask them to go deeper as in:* I heard you mention […] please tell me about other situational factors surrounding that project...]

Probing questions if they aren’t being specific enough, but try to elicit their own factors:
a. Were you working alone or as part of a group? Internal or external?
b. What resources did you have access to? (time, budget, personnel)
c. *Was there a particular champion of your work? A particularly difficult individual to work with?*</td></tr>
<tr><td>6</td><td>Thank you! Now please tell me - even with silly levels of detail - your process for creating the viz. How did you begin?
a. And then what happened next? … [*Repeat as needed.*]
b. At what point did you know you were done?

Probing questions if they aren’t being specific enough, but try to elicit their own process:
a. How did you respond?
b. Could you give me a little bit more detail about…
c. Was there a step between? …
d. What was the purpose of …
e. Resource consideration</td></tr>
</table>

<table>
<tr><td></td><td></td><td>
i. What software did you use? (Tableau, PowerBI, excel, SPSS, R)
ii. How did you choose the software?
iii. Did you use any “non-software” tools in your design (draft on paper/whiteboard, etc.)
iv. What resources did you use to guide your work?
v. How did you manage … (time, personnel, budget?)
f. External
i. Collaboration (who, why, what purpose did the collaboration fulfill? What was the result?)
ii. Research (Did you seek any additional resources during this process?)
iii. Audience: (Same as funders? How early in the process did you think about the audience? Did your knowledge of your audience affect your viz in any way? Did you pilot test, etc.?)
g. Internal
i. Thinking (About the process, of a story, about the audience, anything)
ii. Trying (Multiple attempts, or get it on the first time? Drafts?)
iii. Ways to promote creativity?
</td></tr>
<tr><td colspan="3">Thank you so much for walking through that specific example.</td></tr>
<tr><td rowspan="4">5 min</td><td>7</td><td>Understanding that the above may not be a typical data viz experience, what general process do you recommend?</td></tr>
<tr><td>8</td><td>How would you explain this process to someone who was new to creating data visualization?</td></tr>
<tr><td>9</td><td>What are some challenges you have encountered while developing a data visualization?
Probing questions if they aren’t being specific enough:
a. Do you often have challenges working with your team/ your clients?
b. Does technology ever cause a problem?
c. Do the resources you are given often present a challenge?</td></tr>
<tr><td>10</td><td>I asked before about “why” you add data viz to your work. Do you have any examples of your visualizations being used (brought up, mentioned) by others? [Probe: Is there anything about your process that you believe contributes to the use of this data visualization?]</td></tr>
<tr><td colspan="3">Thank you so much for your time today. As we wrap-up, I’d like to know what advice you have for others when it comes to creating data viz and what questions you would like to ask other data viz experts.</td></tr>
<tr><td rowspan="3">10 min</td><td>11</td><td>What do you wish you had known - or which specific resource do you wish you had had - when you first started doing data viz?</td></tr>
<tr><td>12</td><td>What advice do you have for others when it comes to creating data viz?</td></tr>
<tr><td>13</td><td>What questions would you like to ask other professionals when it comes to creating data viz?</td></tr>
<tr><td colspan="3">Thank you, again, for your time, I really appreciate you sharing your expertise with me. If you have any questions or think of anything else to add, please reach out. And, if you are interested in our results, please let me know and I will certainly follow-up with you after we have reviewed these interviews.</td></tr>
</table>

Appendix D: Examples of Use Provided by Interviewees

Evaluator	#	Example(s)	Interpretation (Evidence) of Use	How or Why Contributed	Discernable Model(s)
AA-10	1	Presentation slide with six improvement suggestions broken down into short-term, medium-term and long-term.	clinic adopted 3 of 6 recommendations	viz clearly articulated easiest to adopt	*Efficiency* *U-FE*
BB-15	1	Visual on school discipline showed a huge [*dramatic*] disparity in African American students that were getting suspended.	*and it went to the board and they looked at it and were like, "Wow."*	*very dramatic compared to just numbers in a table*	*Efficiency* *U-FE* *Data → Insight* *Audience*
CC-30	3	A dashboard for the school board used to track how programs were increasing educational outcomes (academics and wellbeing) for K through 8 students.	grassroots non-profit merged with [a bigger org] and a stated reason for keeping their programs was *how well performance was demonstrated by viz*	our dashboard was a part of that… such a good job of measuring their outcomes	*Efficiency* *U-FE* *Explain ←→Explore* *Audience*
		Providing design (but not data analysis and not build) for 1 page dashboard reports to show the board an annual snapshot of progress for 10 or 15 different environmental projects.	we actually don't know what exactly happened afterwards (lost contact) but the plan was to build in house		*Explain ←→Explore* *Audience*
		[UniversityB] *had multiple departments using one dashboard* and [UniversityA] had eight different departments trying to roll out eight dashboards.	they were reviewed at least annually	informed the development of a trans-disciplinary health initiative	*U-FE* *Explain ←→Explore* *Audience*
EE-4 [a]	2	Started adding nicer data viz and formatting pieces and then all of a sudden [the head person] was very excited about the reports.	*if this is something that's gonna get them so excited then make sure that we are incorporating that on every page*	very *excited* about the reports	*Audience*

Evaluator	#	Example(s)	Interpretation (Evidence) of Use	How or Why Contributed	Discernable Model(s)
		Presentation included icon-based graphic of indicators showing significant relationship to outcomes. …	they really responded well to this … they really *latched on to it* [and showed understanding]	the icon and the way we explained it … helped motivate diving deeper	*U-FE* *Data → Insight* *Audience*
FF-10	1	Dashboard of city park usage data disseminated across the state in a very simple, clear way… something fun appealing to people of different statistical and data backgrounds.	client was very open and receptive to it, so they loved the end product and it was very useful to them	they were able to clearly extrapolate the information that they wanted	*Efficiency* *U-FE* *Explain ←→Explore* *Data → Insight* *Audience*
GG-10	2	Added data viz to the report - trying to create a really exceptional report to justify role of evaluator to very difficult physician clients.	There was discussion about it… ignored the evaluator and "start[ed] doing informed decision-making"	*it was really important that they see the main finding from this chart*	*Efficiency* *U-FE* *Explain ←→Explore* *Audience*
		Visualizations for a provincial report on the impact of COVID on non-profits.	one of our members of government [equivalent to US congressman] actually spoke about the report	easily shared with others	*U-FE* *Audience*
II-6	1	A lot of the work for this project has been presented and disseminated to the relevant local policy holders.	a number of presentations and discussions include these visuals	easy to include in other presentations	*U-FE* *Audience*
JJ-3	0	No example provided.	clients "like" the viz, but does not consider that use		
KK-10	1	Attendance breakdown of youth program. Learned that virtual programs are attracting new kids and that *almost two thirds of them are pretty much coming in once and then they don't come back again, which … isn't something that we really want.*	client repurposed viz: embedded images into emails with provocative messages. Clients indicated *we learned from this*; *elicit them to wanna learn more and do better; creates good conversation*	client can *use it and carry it however they want. They can manipulate it.*	*U-FE* *Data → Insight* *Audience*

Evaluator	#	Example(s)	Interpretation (Evidence) of Use	How or Why Contributed	Discernable Model(s)
LL-4	3	Data visuals within the report were *very much a branding exercise*.	*at least they look through it and it's like, "Oh, this looks pretty. I'm happy with what I paid for."*	pretty visuals	*U-FE (symbolic?)* *Audience*
		Weekly "immediate reports" about student program engagement included a small multiple chart of the 12 students. Intent was to show linear trend for all students but visual sparked client concern about a specific student and asked to break anonymity.	*They're like, "That makes a lot of sense 'cause of this. I will fix it, this reason."* Also viewed asking more questions as "use".	knows intervention was because of visual but indicated is usually the overall reporting	*Efficiency* *U-FE* *Explain ←→Explore* *Data → Insight* *Audience*
		Recreated visuals for HR person to show sub-group variation and not hide variation in the averages.	*Others had seen the first draft of the slides and said, "This doesn't make sense to us. It's a bunch of numbers slashed together."*	*You could visualize the point as opposed to being told it.*	*Efficiency* *U-FE* *Explain ←→Explore* *Data → Insight* *Audience*
MM-7	1	Slides summarizing the big picture of what interviews and focus groups said. Visually, there wasn't a ton to it, it was just kind of distilling the qualitative work and making it tell a story.	*presentation is constantly referenced by our program staff… making a difference on people remembering key takeaways*	*I do think laying it out visually was a key factor*	*Efficiency* *U-FE* *Explain ←→Explore* *Data → Insight* *Audience*
OO-10	1	Data visualization project for a client and they decided to put their organization's annual report on a website as opposed to create a downloadable document.	*So, the visualizations that I created are out there, so being used.*		*U-FE* *Explain ←→Explore* *Audience*
TT-8	3	A report with lots of data visuals. They basically had to stop the program, but still without having served everyone who wanted assistance.	being used and talked about; used to help advocate for more funding and resources	resonated with client: *"Oh, my gosh, that tells exactly the story that we want."*	*U-FE* *Explain ←→Explore* *Audience*

Evaluator	#	Example(s)	Interpretation (Evidence) of Use	How or Why Contributed	Discernable Model(s)
		A beautiful report, but also made one page infographics for school administrators because they wouldn't read a 50-page report.	target admin were provided the visuals		*Efficiency* *U-FE* *Audience*
		A front and back one-pager regarding student successes and challenges.	have taken to community meetings	client *can share whatever slides are appropriate*	*Efficiency* *U-FE* *Audience*
VV-15	1	Multi-slide animation explaining achievement gap was part of a larger presentation during covid. Initially appeared not to be used, but then multiple grant-making organizations asked for follow-ups to get some more clarity on some of the issues.	*it certainly was picked up and utilized and continues to be used by many of the funding entities in the region here;* help them understand topics	*based on the conversations I've had…some people who only latched onto the stuff because they saw it presented in this way*	*U-FE* *Explain ←→Explore* *Data → Insight* *Audience*
WW-6	2	*client that commented on some of the work that we had done for them and they're like... "And one of the things that really sticks out to us is just like the visual of the aesthetics and the time you take to make it really clear and clean." ...I've realized how much that carries over in terms of just pride of work and cleanliness of work.*	*the low-hanging wow factor… you give them these visuals and they're just like, "Oh my gosh, what is this?"*	aesthetics and "low-hanging wow factor"	*Efficiency* *Audience*
		Created simple Tableau dashboards visualizing COVID data and trained the clients. Inspired client to start using and engaging in Tableau and continue building out the dashboard.	*help empower [client] to start using the tool and now they're doing all these really cool analyses in Tableau*	inspire clients to do more with data	*U-FE* *Explain ←→Explore* *Data → Insight* *Audience*
XX-8 [b]	1	Interactive dashboard of Covid-19 vaccine acceptance data to help policy makers plan health communications campaigns with 18 longitudinal survey waves and 67 countries involved.	widely used; feedback from key audiences has been positive in terms of having access to the data	access to data caused client to ask more questions and want further analysis	*U-FE* *Explain ←→Explore* *Data → Insight* *Audience*

Evaluator	#	Example(s)	Interpretation (Evidence) of Use	How or Why Contributed	Discernable Model(s)
YY-5	4	Reporting dashboard commissioned by university, but requires reporting from local site teams. Reporting of data has decreased.	this indicates the site teams are not “using”		*U-FE* *Audience*
		Conducted a needs assessment and built data visualizations and dashboards used to facilitate a process to select their intervention	emphasized clients questioning the data and connected ownership as engagement	our dashboards were used to help inform the decision	*U-FE* *Explain ←→Explore* *Data → Insight* *Audience*
		Dashboards of training projects regarding feedback on the quality and the impact of trainings.	views sparking dialog and getting feedback as evidence of use	*giving the target audience what they want in terms of format of the data (website, slide deck, handout, etc.)*	*U-FE* *Audience*
		Dashboards for an online mandated reporter project detailing usage levels and efficiencies they were hoping to gain. Also collected and provided feedback on the user experience.	feedback informed improvement decisions and revisions to the agency's website		*U-FE* *Explain ←→Explore* *Data → Insight* *Audience*
ZZ-10 [a]	1	Presentation slides with icons also converted to printed one pager because knew client would print and hand it to trainers and other stakeholders.	one pagers have been handed out to trainers	provided in format that can pass along to someone else; "useful" because they get to take credit for it	*U-FE* *Audience*

Note. Interviews marked [a] were not asked the specific question about evaluation use. The interview marked [b] was not asked the specific question about evaluation use, but likely because the entire interview centered around use. Direct quotes are presented in italics.

Appendix E: Follow-up Interview Questions

<table>
<tr><td></td><td>Introduction and consent.</td></tr>
<tr><td>5 min</td><td>First, I'd like to thank you – again – for taking the time to meet with me. I cannot express how much I appreciate your help with my project. So, thank you!

Second, going to explain the background for this discussion and then ask for your consent – again – to be interviewed.

I'm talking with you today because you are both an evaluator and an expert in data visualization. Our research team interviewed you two years ago about your process for developing data visualizations. We presented those findings at AEA in 2021 and 2022 and we have an article discussing data viz challenges under review with Eval and Program Planning. So, thank you so much for your time then --- and now!

I asked you to meet with me again because I need your help exploring a theme that emerged in the first round of interviews and needs clarification. In the first interview, we asked evaluators WHY they use data viz in their work. It sounds like a simple and straightforward question but what emerged from the answers is a conceptual model of evaluators using data viz for audience engagement -- meaning that evaluators spoke at length and in detail about using data viz as a way to engage audiences. Some subthemes of audience engagement were very clear within the discussions --- but the way(s) in which those subthemes are organized and interact is not clear from the discussions. This is because that discussion of audience engagement was not the point of the original interview so no follow-up questions were asked. We didn't even know it existed until we went back through the interview data. Therefore, I still have some gray areas within the model that I would really appreciate discussing with you today.

My plan is to areas of this emerging model with you and ask you how you think about using data viz for audience engagement within evaluation. Does that work for you?

So, the other thing I need to ask you is to keep our discussion confidential – not to even mention that there is an emerging model of audience engagement specific to evaluation – until after I get a chance to test it with other evaluators and share my research formally. Basically, even telling others that someone is working on a model of audience engagement in eval data viz would be a spoiler and I don't want any spoilers. Is that okay with you?

Thank you! And lastly, I just want to confirm that you are willingly participating in this interview and you consent to it being recorded so that I can transcribe it.

Thank you!</td></tr>
<tr><td></td><td>Overview.</td></tr>
<tr><td>15 min</td><td>Before we jump into specifics, I want you to tell me a bit about how you think about data visualization's role in engaging an audience within evaluation.

What does audience mean to you? [Do you prefer the term audience or stakeholder? Why?]

What does audience engagement mean to you?

Have you ever specifically thought about data visualization as a way to engage an audience?

Who is typically the audience for your viz?

Is it important to engage an audience? Why? What happens if you don't engage the audience?</td></tr>
</table>

<table>
<tr><td></td><td>How and why do you think data viz engages an audience?

What can you do with your data viz to make it better (or worse) at engaging an audience?</td></tr>
<tr><td></td><td>Detailing subthemes. Reminder for self: Not to share with interviewees – at least not in initial interviews.


</td></tr>
<tr><td>30 min</td><td>Thank you! Now I would like to talk about specific topics related to audience engagement and get your insight into how you think about these things. [Address areas that the specific participant is best able to assist with. Not all interviews will discuss all subthemes. New subthemes may emerge or become merged in subsequent interviews.]

Satisfaction

What do you think is the role of data viz in increasing audience satisfaction with evaluation?
How do you know if clients are satisfied with the data viz you provide?
What do you do to promote client satisfaction with your data viz?
What is the relationship between quality of data viz, client satisfaction, and clients USING your data viz?
Have you ever had clients express satisfaction or dissatisfaction with your data viz? How did that reaction change your approach?
Can you think of a scenario where the client’s satisfaction with your data viz doesn’t matter?

Capacity building

What does evaluation capacity building mean to you?
What do you think about capacity building in the context of data viz within evaluation?
How involved are your clients in the data viz process?

Does your work include helping clients interpret data viz? Create their own data viz? Tell me about that.
Do you show clients drafts of your data viz and get their feedback?
Does it matter if they participate in data viz creation? Why?

If you show clients how to create their own data viz, are you working yourself out of a job?

Brand identity

Do you incorporate client brand image and style-guides into your data viz work? If so, how and why?
What would happen if you didn’t use their branding?
What do you do when client brands and style-guides are not compliant with design needs? [not enough color contrast, not color-blind appropriate, etc.]
</td></tr>
</table>

<table>
<tr><td></td><td>
Have you experienced your data viz being used for client marketing or promotion purposes? What is your response to this? [is it appropriate? Seeking to understand their line between evaluator and marketer.]
Provide artifacts

How do you typically provide data viz to clients? In a formal report, in a presentation?
Why do you do it this way?
Do you find that clients want their data viz “a certain way” or are they expecting you to determine the format and deliverable?
How do you know the format in which clients want their data viz?
Do you find that your data viz gets used in other ways than you expected? What are some examples?
Have you ever had your data viz used inappropriately?

Additional audiences

Do you find your clients using your data viz to share with others? Give me some examples of that.
Do you do anything special with your data viz to help it be used in different ways?
What is the most unexpected thing you’ve had happen with your data viz? Clients using it in an unexpected way?

</td></tr>
<tr><td></td><td>Model validation: If interaction between sub-themes is sufficiently developed, then later interviewees may be presented with the most recent version of the model and asked for specific feedback.
Thank you! I would like to share my screen and quickly walk you through some of the sub-themes within the model as I understand them now. My goal is to get your insight into how you think about these things.
[Share and describe model.]</td></tr>
<tr><td></td><td>Does it ring true to you?

What parts make the most sense to you?
Are these the major components of audience engagement? What is missing? What is extra?

How would you improve it?

What is missing?
Is any of this “extra” and should be removed?

You’re a data viz expert :), how would you draw this model – meaning would you visualize it differently than I am presenting it here?</td></tr>
<tr><td></td><td>Conclusion and thank you</td></tr>
<tr><td>5 min</td><td>Thank you, again, for your time, I really appreciate you sharing your expertise with me. If you have any questions or think of anything else to add, please reach out. And, if you are interested in our results, please let me know and I will certainly follow-up with you.</td></tr>
</table>

Appendix F: Follow-up Interview Participant Communication

Participant invitation email:

Subject line: Data Visualization Process (Follow-up Interview Request)

Dear [xxxxx],

You participated in an interview 2 years ago to help our data viz research lab better understand your experiences and processes creating data visualizations.

We presented findings from that data at AEA 2021 and AEA 2022. (Thank you!)

Would you be willing to participate in a follow-up interview? There was one finding from the first round of data collection that we want to investigate further. The topic we are investigating concerns **how evaluators consider audience in their data viz design process**.

If you are willing to explore this data viz audience topic further, please review and complete the consent form here [link to: https://cgu.co1.qualtrics.com/jfe/form/SV_aa93oimB9zeW8fk] and hit submit. I will then contact you to set up an interview time.

If you do not wish to participate, please let me know!

Sincerely,

Sarah Douville, PhD Student
Claremont Graduate University
Data Visualization Lab Researcher
Sarah.Douville@cgu.edu
(909) 201-6617

(Yes, consent form completed) Participant interview invitation email:

Subject line: Data Visualization Process (Schedule a Follow-up Interview)

Dear [xxxxx],

Thank you for agreeing to participate in a follow-up interview. I expect this interview to last 30 - 60 minutes, so please let me know if any of the times below work for you and I will send a calendar invite with zoom link.

All times are stated in California time. If these times do not work for you, I will provide additional options – just let me know!
[insert availability for following two weeks]

Sincerely,
Signature Lines

Appendix G: Follow-up Interview Informed Consent

Claremont Graduate University

AGREEMENT TO PARTICIPATE IN DATA VISUALIZATION FOLLOW-UP INTERVIEW(S) (IRB # 4579)

You are invited to continue your participation in a research study. Volunteering will probably not benefit you directly, but you will be helping us better understand one aspect of data visualization within program evaluation. If you volunteer, you will be asked to participate in a 30-60 minute interview and possibly additional interviews or email exchanges if questions remain. This will take about 60-90 minutes of your time. Volunteering for this study involves no more risk than what a typical person experiences on a regular day. Your involvement is entirely up to you. You may withdraw at any time for any reason. Please continue reading for more information about the study.

Study Leadership: This research study is led by Sarah Douville, a graduate student of Claremont Graduate University, who is being supervised by Tarek Azzam, a professor at Claremont Graduate University with aid from the data visualization lab.

Purpose: The purpose of this study is to better understand how program evaluators think about the process of creating a data visualization.

Eligibility: To be in this study you must be a program evaluator, create data visualization professionally, and be able to discuss the process as you have experienced it.

Participation: During the study, you will be asked to participate in one interview conducted by Sarah Douville. Data from this study will be analyzed and reported in academic journal(s) in an attempt to further the field of data visualization. Participation will take between 60-90 minutes and questions will ask about your experience and process for creating data visualization.

Risks Of Participation: The risks you run by taking part in this study are minimal and are not greater than those ordinarily encountered in daily life or during the performance of routine work.

Benefits Of Participation: We **do not** expect the study to benefit you personally. This study is intended to benefit the data visualization community by informing them as to common practices in data visualization.

Compensation: You will not be directly compensated for participating in this study. Your participation is greatly appreciated.

Voluntary Participation: Your participation in this study is completely voluntary. You may stop or withdraw from the study at any time without it being held against you. You may also refuse to answer any question at any time. Your decision whether or not to participate will have no effect on your current or future connection with anyone at CGU.

Confidentiality: Your individual privacy will be protected in all papers, books, talks, posts, or stories resulting from this study. We may share the data we collect with other researchers, but we will not reveal your identity with it. In order to protect the confidentiality of your responses, we will assign each participant a pseudonym. Files containing your true identity, such as your consent form and audio/video recordings will be stored in password protected accounts accessed only by the research team.

At the beginning of the interview we will ask to digitally record the conversation. These digital recordings will be saved from the recording device to a password protected shared drive immediately after the interview is completed. After checking to ensure the transition was successful, the recording will be deleted from the digital recording device.

The digital recordings will be used to produce transcripts/detailed notes for analysis. If at any point your name is mentioned during the interview, it will simply be replaced with "the interviewee" or your pseudonym when transcribed. Audio/video recordings will be destroyed after transcripts have been created and verified. Transcripts will be identified only by pseudonym and will be stored in password protected accounts accessed only by the research team.

Survey data and transcripts/detailed notes will be retained by the researchers until publications from the research study have been developed and presented in the peer-reviewed literature.

Further Information: If you have any questions or would like additional information about this study, please contact Sarah Douville at sarah.douville@cgu.edu, (909) 210-6617 or Tarek Azzam at tarek.azzam@cgu.edu. The CGU Institutional Review Board has certified this project as exempt. If you have any ethical concerns about this project or about your rights as a human subject in research, you may contact the CGU IRB at (909) 607-9406 or at irb@cgu.edu. You may print and keep a copy of this consent form. **OR** If you wish, we will be happy to send you a copy of this consent form.

Consent: Typing your name below, marking "yes," and clicking "submit," indicates that you understand the information on this form, that someone has answered any and all questions you may have about this study, and you voluntarily agree to participate in it. If you do not consent, you may select the option to not participate or simply leave this page without providing your name. Thanks!

____

Qualtrics survey options:

- **Yes**. I consent to participate in this data visualization research project. Please contact me for a follow-up interview.
- **No**. I do NOT consent. I will NOT participate.

Appendix H: Survey Participant Communication

Subject line: Survey: How do you think about using Data Viz in Evaluation?

Dear [xxxxx],

I am conducting a survey study to better understand why program evaluators use data visualization (data viz) and the conceptual models they use when thinking about data viz use. If you have other colleagues (including students) in evaluation who may wish to participate, please forward this email.

What you will be asked to do:

You will be presented with short explainer videos describing different ways of conceptualizing data viz use in evaluation. You will then be asked questions about your familiarity with that model and how useful you find it. No deception is used in this study and the models presented to you are based on evidence and research from evaluation and other fields. Therefore, you may learn something useful about data viz just for participating.

Participation in this anonymous survey will take approximately 20 minutes.

To participate in the survey:

More information and a consent form are included in the survey here [link to: https://cgu.co1.qualtrics.com/xxxxx **<-unique link for each list**]

If you do not wish to participate, please disregard this email.

Sincerely,

Sarah Douville, PhD Student
Claremont Graduate University
Data Visualization Lab Researcher
Sarah.Douville@cgu.edu
(909) 201-6617

Appendix I: Survey Informed Consent

Start of Block: Consent

Consent **Introduction:**

As a graduate student at Claremont Graduate University, I am conducting a study to better understand why program evaluators use data visualization (data viz) and the conceptual models they use when thinking about data viz use.

What you will be asked to do:

You will be presented with short explainer videos describing different ways of conceptualizing data viz use in evaluation. You will then be asked questions about your familiarity with that model and how useful you find it. No deception is used in this study and the models presented to you are based on evidence and research from evaluation and other fields. Therefore, you may learn something useful about data viz just for participating.

Please do not talk about or share the video models presented until I have had a chance to formally share my findings.
Please do not take this survey more than one time.

Participation in this anonymous survey will take approximately 20 minutes.

Consent:
STUDY LEADERSHIP. You are being asked to take part in a research project that is led by Sarah Douville, a graduate student at Claremont Graduate University, who is being supervised by Professor of Evaluation and Applied Methods, Tarek Azzam.
PURPOSE. The purpose of this study is to better understand conceptualizations of data viz use within program evaluation.
ELIGIBILITY. To be in this study, you must be an evaluator. Non-evaluators are welcome to view the videos and contribute their data, but they are not the primary focus of this research.
PARTICIPATION. During the study, you will be asked to provide information about your experience in evaluation and data viz, view a brief introductory video on data viz and answer questions on this, then you will view another brief video describing one way of conceptualizing data viz use and answer questions on this video. If you would like to continue to explore other ways of conceptualizing data viz use, you will be provided additional video(s) and questions. Finally, you will be asked a few demographic questions. This should take approximately 20 minutes to complete.
RISKS OF PARTICIPATION. The risks that you run by taking part in this study are minimal.
BENEFITS OF PARTICIPATION. You will not receive any money from participating, but you may learn new ways of thinking about data viz in your evaluation work. This study is intended to benefit the field of evaluation by contributing to the knowledge base on data viz use. This study will also benefit the researcher by helping me complete my graduate education.
COMPENSATION. There is no direct compensation to you for participating in this study.
VOLUNTARY PARTICIPATION. Your participation in this study is completely voluntary. You may stop or withdraw from the study at any time or refuse to answer any particular question for any reason without it being held against you. Your decision whether or not to participate will have no effect on your current or future connection with anyone at CGU.
CONFIDENTIALITY. Your participation in this survey is anonymous. The survey asks for your demographic data, but does not request your name, email, or other personally identifying information.
FURTHER INFORMATION. If you have any questions or would like additional information about this study, please contact: Sarah Douville, Sarah.Douville@cgu.edu, (909) 210-6617. You may also contact my faculty advisor: Tarek Azzam, TarekAzzam@UCSB.edu, (909) 374-5355.
The CGU Institutional Review Board has certified this project exempt from IRB oversight [CGU IRB #4655]. You may print and keep a copy of this consent form.

Consent: Continue By selecting "continue" below, you agree to participate in this survey.

○ Continue (1)

○ I choose not to participate (2)

End of Block: Consent

Appendix J: Survey Questions

Start of Block: Identification as an evaluator

Q41 Anyone is welcome to watch these videos and provide their data, but this study is concerned with how evaluators consider data visualization in their work.

Are you an evaluator? (This includes any type of evaluation such as program, product, policy, assessment, etc. and in any capacity such as student, faculty, or practicing evaluator.)

- ○ Yes, I am an evaluator (1)
- ○ No, I am not part of this profession (2)

End of Block: Identification as an evaluator

Start of Block: Experience (Evaluators) [These questions only provided to Evaluators]

Please let me know more about your experience with evaluation.

Which of the following best describe your current professional identity in the evaluation field? (check all that apply)

- ▢ Evaluator (in any capacity) (1)
- ▢ Student involved in evaluation (paid or unpaid) (2)
- ▢ College or university faculty member or instructor (3)
- ▢ Researcher (4)
- ▢ Retired, but still active in the evaluation field (5)
- ▢ Trainer (6)
- ▢ Unemployed (7)
- ▢ I don't identify myself with the evaluation field (8)
- ▢ Other: (9) ____

How may years have you been associated with the evaluation profession?

____

Are you engaged in evaluation:

- ○ Full-time (1)
- ○ Part-time (2)
- ○ Not at all (3)

What is your typical role in an evaluation?

- ○ Internal evaluator (1)
- ○ External evaluator (2)
- ○ Mixed internal/external evaluator (3)
- ○ Other (5) ____

Have you ever been a member of a formal evaluation association? (AEA, CEA, local affiliate, etc.)

- ○ Yes (current member) (1)
- ○ Yes (past member) (2)
- ○ No (3)

How would you describe your evaluation knowledge and skills?

- ○ Expert (1)
- ○ Advanced (2)
- ○ Intermediate (3)
- ○ Novice (4)
- ○ Very limited or none (5)

Approximately how many evaluations have you completed?

____

Please describe your personal approach to evaluation.

____
____
____

Page Break

Please let me know more about your experience with data visualization.

The definition of data viz used in this study is:

"...a process that (a) is based on qualitative or quantitative data and (b) results in an image that is representative of the raw data, which is (c) readable by viewers and supports exploration, examination, and communication of the data" (Azzam, Evergreen, Germuth, and Kistler, 2012, p. 9.)

This research will restrict the definition of data viz to the visual display of **data** and will not address reasons to focus on formatting and graphic design in evaluation.

How often do you use data visualization in your evaluation work?

- ○ Always (4)
- ○ Usually (5)
- ○ Sometimes (6)
- ○ Rarely (7)
- ○ Never (8)

How would you describe your data visualization knowledge and skills?

- ○ Expert (1)
- ○ Advanced (2)
- ○ Intermediate (3)
- ○ Novice (4)
- ○ Very limited or none (5)

In general, how easy or hard do you find it to...

	Very Easy (1)	Easy (2)	Neutral (3)	Hard (4)	Very Hard (5)
Understand statistics	○	○	○	○	○
Understand graphs or charts	○	○	○	○	○
Create graphs or charts	○	○	○	○	○

In a typical month, how often do you...

	Never (1)	Less Than Once a Month (2)	Once a Month (3)	Once a Week (4)	Once a Day or More (5)
see data displayed in a chart?	○	○	○	○	○
use data displayed in a chart to make decisions?	○	○	○	○	○
create a chart to display data?	○	○	○	○	○
learn new ways of displaying data?	○	○	○	○	○

End of Block: Experience (Evaluators)

Start of Block: Efficiency [All Participants watch this video first and answer 3 questions]

Watch Efficiency Please watch this 3 minute video on data viz efficiency.
Continue the survey after the video is finished.

Indicate your level of agreement with the following statements:

	Strongly Disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly Agree (5)
Data visualization can be used to increase efficiency over text-based communication.	○	○	○	○	○
Before I watched the video, I was familiar with the efficiency rationale for data viz.	○	○	○	○	○
The efficiency rationale for data viz makes sense to me.	○	○	○	○	○

End of Block: Efficiency

Start of Block: Model: Either Explain or Explore, Data to Insight, or Audience Engagement **[Participant is provided 1 video at random and asked the following questions]**

Please watch this video describing one way to think about data viz in evaluation.
Continue the survey after watching the video.

Indicate your level of agreement with the following statements:

	Strongly Disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly Agree (5)
Before I watched the video, I was familiar with the *Audience Engagement* way of thinking about data viz.	○	○	○	○	○
I have learned something new about the *Audience Engagement* model from watching this video.	○	○	○	○	○
I have used the *Audience Engagement* model to describe my reason for using data viz to others.	○	○	○	○	○
I am likely to use the *Audience Engagement* way of thinking about data viz in future evaluation work.	○	○	○	○	○

Indicate your level of agreement with the following statements:

The *Audience Engagement* way of thinking about data viz...

	Strongly Disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly Agree (5)
makes sense to me.	○	○	○	○	○
is appropriate in evaluation.	○	○	○	○	○
aligns with my personal approach to evaluation.	○	○	○	○	○
would benefit stakeholders.	○	○	○	○	○
adds value to the "efficiency rationale" for using data viz.	○	○	○	○	○

I have used the *Audience Engagement* model in my evaluation work.

- ○ always (4)
- ○ often (5)
- ○ sometimes (6)
- ○ a few times (7)
- ○ never (8)

If you have used it, please describe one or two outcomes from using the Audience Engagement model in your evaluation work.

____

____

I think the strengths of the Audience Engagement model are:

____

____

I think the weaknesses of the Audience Engagement model are:

____

____

End of Block: Model: Audience Engagement

[Participant is thanked and offered an opportunity to exit the survey or watch another video at random.]

Start of Block: Continue1? Do you wish to watch another video?

Thank you for your feedback on this data viz model! You may end the survey now. Or, if you *wish* to watch another video and answer another set of questions, your additional feedback would be greatly appreciated.

- ○ I am done watching videos (1)
- ○ I would like to watch another video (2)

End of Block: Continue1? Do you wish to watch another video?

[If participant exits, they are taken to the demographic questions. If participant continues, they watch another video and answer the same questions PLUS these two additional questions that are customized based on the videos watched so far.]

Start of Block: 2Qs (A2) Compare Models

Considering the models you have learned about during this survey, which model is most useful to your evaluation work?

- ○ *Explain or Explore* (1)
- ○ *Data to Insight* (2)
- ○ *Audience Engagement* (4)

Why did you select that model as most useful?

____

____

Start of Block: Demographics (E and NE) **[Demographics are requested after participant watches model videos.]**

Thank you for participating!

To ensure transparency about the various backgrounds and perspectives represented among those who have participated in this survey, please provide more information about your personal background by answering the following questions.

Gender:

- ○ Male (1)
- ○ Female (2)
- ○ Self identify (3) ____
- ○ Choose not to respond (4)

Race:

- ○ American Indian or Alaska Native (1)
- ○ Asian (2)
- ○ Black or African American (3)
- ○ Native Hawaiian or Pacific Islander (4)
- ○ White (5)
- ○ Self Identify (6) ____

Ethnicity:

- ○ Hispanic or Latino (1)
- ○ Not Hispanic or Latino (2)
- ○ Self Identify (3) ____

How do you rate your English language skill?

- ○ Beginner (1)
- ○ Intermediate (2)
- ○ Advanced (3)
- ○ Fluent/Native Speaker (4)

Year of birth:

____

Highest degree received:

- ○ Doctorate (1)
- ○ Masters (2)
- ○ Bachelors (3)
- ○ Other: (4) ____

Which of the following best describes your primary work setting?

- ○ College/University (1)
- ○ School system (2)
- ○ State agency (3)
- ○ Federal agency (4)
- ○ Local agency (5)
- ○ Private business (6)
- ○ Non-profit organization (7)
- ○ Indigenous nation government or governmental organization (8)
- ○ Other: (9) ____

End of Block: Demographics (E and NE)

Appendix K: Phase III Descriptive Statistics for 132 Participants

							Skewness		Kurtosis		Shapiro-Wilk	
	N	Missing	Mean	SD	Min	Max	Skew	SE	Kurt	SE	W	p
Demo_age	109	23	47.05	13.83	23	78	0.46	0.23	-0.81	0.46	0.95	< .001
Exp_years	132	0	15.06	10.41	0	50	0.95	0.21	0.69	0.42	0.93	< .001
Exp_#evals	122	10	45.59	52.62	0	300	2.46	0.22	7.77	0.44	0.74	< .001
KS_eval	132	0	3.92	0.72	2	5	-0.27	0.21	-0.10	0.42	0.83	< .001
KS_dataviz	132	0	3.42	0.74	1	5	0.04	0.21	0.36	0.42	0.84	< .001
dataviz_HowOftenUse	132	0	4.19	0.84	2	5	-0.76	0.21	-0.15	0.42	0.81	< .001
KS_understand-statistics	132	0	3.8	0.91	2	5	-0.52	0.21	-0.40	0.42	0.85	< .001
KS_understand-graphs	132	0	4.36	0.71	1	5	-1.30	0.21	3.17	0.42	0.74	< .001
KS_create-graphs	132	0	3.98	0.88	1	5	-0.78	0.21	0.43	0.42	0.84	< .001
Experience_see-chart	131	1	4.4	0.78	1	5	-1.52	0.21	2.92	0.42	0.73	< .001
Experience_use-chart	132	0	3.51	0.93	2	5	-0.08	0.21	-0.83	0.42	0.88	< .001
Experience_create-chart	131	1	3.53	0.96	1	5	-0.34	0.21	-0.45	0.42	0.89	< .001
Experience_learn-dataviz	131	1	2.69	0.80	1	5	0.97	0.21	0.78	0.42	0.80	< .001
Efficiency_increased	132	0	4.75	0.54	1	5	-3.27	0.21	16.67	0.42	0.49	< .001
Efficiency_familiar-before	132	0	4.09	0.98	1	5	-1.08	0.21	0.51	0.42	0.79	< .001
Efficiency_makes-sense	132	0	4.62	0.66	1	5	-2.32	0.21	7.62	0.42	0.60	< .001
FIRST_familiar-before	132	0	3.36	1.26	1	5	-0.41	0.21	-1.04	0.42	0.87	< .001
FIRST_learned-new	132	0	3.93	0.88	1	5	-0.98	0.21	1.22	0.42	0.82	< .001
FIRST_described-others	132	0	2.64	1.25	1	5	0.35	0.21	-0.99	0.42	0.89	< .001
FIRST_likely-to-use	132	0	3.99	0.90	1	5	-0.89	0.21	0.89	0.42	0.84	< .001
FIRST_makes-sense	132	0	4.47	0.68	1	5	-1.50	0.21	3.98	0.42	0.70	< .001
FIRST_appropriate-eval	132	0	4.48	0.61	3	5	-0.76	0.21	-0.39	0.42	0.72	< .001
FIRST_aligns-approach	132	0	4.26	0.83	1	5	-1.17	0.21	1.99	0.42	0.77	< .001
FIRST_benefit-stakehldrs	132	0	4.34	0.76	1	5	-1.19	0.21	2.03	0.42	0.76	< .001
FIRST_adds-value-efficy	132	0	4.22	0.85	1	5	-0.97	0.21	0.74	0.42	0.80	< .001
FIRST_have-used	130	2	2.68	1.40	1	5	-0.07	0.21	-1.54	0.42	0.82	< .001

Note. Red indicates cut-offs were violated.

Appendix L: Phase III Descriptive Statistics for 131 Participants

	N	Missing	Mean	SD	Min	Max	Skewness		Kurtosis	
							Skew	SE	Kurt	SE
Demo_age	108	22	46.83	13.72	23	78	0.480	0.233	-0.762	0.461
Exp_years	131	0	14.84	10.11	0	50	0.906	0.212	0.653	0.420
Exp_#evals	121	10	45.55	52.84	0	300	2.448	0.220	7.690	0.437
KS_eval	131	0	3.92	0.72	2	5	-0.263	0.212	-0.125	0.420
KS_dataviz	131	0	3.42	0.74	1	5	0.052	0.212	0.360	0.420
dataviz_HowOftenUse	131	0	4.20	0.84	2	5	-0.791	0.212	-0.080	0.420
KS_understand-statistics	131	0	3.79	0.91	2	5	-0.514	0.212	-0.421	0.420
KS_understand-graphs	131	0	4.37	0.71	1	5	-1.345	0.212	3.458	0.420
KS_create-graphs	131	0	3.98	0.88	1	5	-0.777	0.212	0.437	0.420
Experience_see-chart	130	1	4.40	0.78	1	5	-1.528	0.212	2.919	0.422
Experience_use-chart	131	0	3.52	0.92	2	5	-0.087	0.212	-0.809	0.420
Experience_create-chart	131	1	3.53	0.96	1	5	-0.339	0.212	-0.454	0.420
Experience_learn-dataviz	131	1	2.69	0.80	1	5	0.973	0.212	0.784	0.420
Efficiency_increased	131	0	4.78	0.44	3	5	-1.639	0.212	1.476	0.420
Efficiency_familiar-before	131	0	4.09	0.98	1	5	-1.082	0.212	0.486	0.420
Efficiency_makes-sense	131	0	4.65	0.58	2	5	-1.680	0.212	3.046	0.420
FIRST_familiar-before	131	0	3.35	1.25	1	5	-0.407	0.212	-1.042	0.420
FIRST_learned-new	131	0	3.95	0.84	1	5	-0.862	0.212	0.924	0.420
FIRST_described-others	131	0	2.66	1.25	1	5	0.340	0.212	-0.993	0.420
FIRST_likely-to-use	131	0	4.02	0.86	1	5	-0.768	0.212	0.531	0.420
FIRST_makes-sense	131	0	4.50	0.61	3	5	-0.802	0.212	-0.331	0.420
FIRST_appropriate-eval	131	0	4.49	0.61	3	5	-0.773	0.212	-0.370	0.420
FIRST_aligns-approach	131	0	4.28	0.78	1	5	-0.940	0.212	1.043	0.420
FIRST_benefit-stakehldrs	131	0	4.37	0.70	2	5	-0.789	0.212	-0.071	0.420
FIRST_adds-value-efficy	131	0	4.24	0.81	2	5	-0.744	0.212	-0.268	0.420
FIRST_have-used	129	2	2.69	1.40	1	5	-0.087	0.213	-1.528	0.423

Note. Red indicates cut-offs were violated.

Appendix M: Study Assumptions of Random Assignment and Balanced Design

<table>
<tr><th>Assumption</th><th>Met?</th><th>Rationale</th></tr>
<tr><td>Random Assignment</td><td>Yes</td><td>Participant assignment to video watch order was randomized through Qualtrics and double-checked via chi-square tests of independence. Self-identify race responses were collapsed into one category to reduce cells with counts less than five. Many variables still had cell counts less than five. All categorical variables were non-significant:
<table>
<tr><th>Variable</th><th>X^2</th><th>df</th><th>p</th></tr>
<tr><td>Gender</td><td>2.21</td><td>6</td><td>0.900</td></tr>
<tr><td>Race</td><td>10.95</td><td>8</td><td>0.205</td></tr>
<tr><td>English Fluency</td><td>7.90</td><td>6</td><td>0.245</td></tr>
<tr><td>Education</td><td>9.52</td><td>6</td><td>0.146</td></tr>
<tr><td>Work Setting</td><td>18.90</td><td>18</td><td>0.398</td></tr>
<tr><td>Professional ID</td><td>5.39</td><td>6</td><td>0.495</td></tr>
<tr><td>Full-time/Part-time</td><td>1.83</td><td>4</td><td>0.766</td></tr>
<tr><td>Role</td><td>9.75</td><td>6</td><td>0.136</td></tr>
</table></td></tr>
<tr><td>Balanced Design</td><td>Yes</td><td>There is no significant difference between the number of people assigned to each IV group and this assignment was random.

Of the 131 participants, 42 (32.1%) were presented the explain ←→ explore model, 43 (32.8%) the data → insight model, and 46 (35.1%) the audience engagement model video. A chi-square goodness of fit calculation confirms that the number of participants randomly assigned to each of these three conditions are equal as the difference between each and the expected value of 43.7 is not significantly different ($X^2(2) = 0.20$, $p = .906$).</td></tr>
</table>

Appendix N: Chi-Square Test Assumptions

Chi-Square Goodness-of-Fit Test Assumptions

		Rationale	
Assumption	Met?	Balanced Design Question	Model Preference Question
#1: One categorical variable	Yes	Each model (EE, DI, AE) is a categorical group.	Each model (EE, DI, AE) is a categorical group.
#2: Independence of observations	Yes	Analysis will only consider the first model viewed so that no participant is in more than one group.	Participants can select only one of the three models (EE, DI, AE).
#3: All cells have expected counts greater than five	Yes	There are 131 observations for 3 possible categories; the expected values of 43.67 are all greater than 5.	24 participants selected one of three models (EE, DI, AE); the expected values of 8 are greater than 5.

Chi-Square Test for Association Assumptions

Assumption	Met?	Rationale
#1: Two categorical variables	Yes	Each model (EE, DI, AE) is a categorical group and the decision to exit or continue is categorical.
#2: Independence of observations	Yes	Analysis will only consider the first model viewed so that no participant is in more than one group.
#3: All cells have expected counts greater than five	Yes	There are 2 possible outcomes for each model; the expected values of 21, 21, 21.5, 21.5, 23, and 23 are all greater than 5.

Appendix O: ANOVA Assumptions

Assumption	Met?	Rationale
#1: One continuous DV	Yes	The 10 model questions have responses on a 5-point Likert-type scale that will be treated as continuous, consistent with the discussion above.
#2: One IV of two or more categorical, independent groups.	Yes	Each model (EE, DI, AE) is a categorical group and no person will be placed in more than one group for ANOVA analysis.
#3: Independence of observations	Yes	ANOVA analysis of the 10 model questions will only consider the first model viewed. This will ensure that no participant is in more than one group.
#4: No significant outliers in IV groups by DV	No	Reviewing box plots divided by IV suggests outliers in the data. These outliers generally correspond to the few very low scores displayed in Figure 28. I reviewed the data again and still believe the outlier responses are their intentional answers and valid data. In essence, if there are mean differences between groups, these outliers are likely the cause. I chose to proceed on the common wisdom that ANOVA is robust when sample sizes are equal.
#5: DV should be approximately normally distributed for each group	No	Responses to the 10 model questions were analyzed by group, see Appendix P for details. Shapiro-Wilkes continued to be significant at $p < .001$ for all variables. The skew cut-off of +/-1 was violated for 3 questions in the EE model and 3 different questions in the DI model. The Kurtosis cut-off of +/-3 was violated for “aligns with personal approach” in the DI model. With the highest skew violation at 1.63 and the single kurtosis violation at 3.76, I determined that the violations were not extreme enough to negate the common wisdom that ANOVA is robust when sample sizes are equal. Therefore, I chose to proceed.
#6: Homogeneity of variances in each group	No	The assumption of homogeneity of variances was violated, as assessed by Levene's test for equality of variances for two questions: familiar before watching ($p < .001$) and makes sense ($p = .003$). I will, therefore, report Welch ANOVA and results from Games-Howell post hoc test for model comparisons, as recommended by Laerd Statistics (2024).

Appendix P: Phase III Descriptive Statistics for 131 Participants by Model (First View Only)

Model								Skewness		Kurtosis	
	Question	N	Missing	Mean	SD	Min	Max	Skew	SE	Kurt	SE
EE	familiar-before	131	0	2.74	1.449	1	5	0.231	0.365	-1.402	0.717
	learned-new	131	0	3.93	0.867	1	5	-1.039	0.365	2.102	0.717
	described-others	131	0	2.31	1.158	1	5	0.440	0.365	-0.878	0.717
	likely-to-use	131	0	4.14	0.899	1	5	-1.351	0.365	2.636	0.717
	makes-sense	131	0	4.50	0.741	3	5	-1.134	0.365	-0.174	0.717
	appropriate-eval	131	0	4.60	0.544	3	5	-0.876	0.365	-0.306	0.717
	aligns-approach	131	0	4.31	0.749	3	5	-0.585	0.365	-0.961	0.717
	benefit-stakehldrs	131	0	4.33	0.786	2	5	-0.996	0.365	0.449	0.717
	adds-value-efficy	131	0	4.21	0.842	2	5	-0.691	0.365	-0.497	0.717
	have-used	130	1	2.50	1.436	1	5	0.130	0.365	-1.666	0.717
DI	familiar-before	131	0	3.70	1.059	2	5	-0.487	0.361	-0.933	0.709
	learned-new	131	0	4.07	0.828	2	5	-0.926	0.361	0.867	0.709
	described-others	131	0	2.86	1.283	1	5	0.131	0.361	-1.079	0.709
	likely-to-use	131	0	3.95	0.844	2	5	-0.658	0.361	0.168	0.709
	makes-sense	131	0	4.65	0.482	4	5	-0.657	0.361	-1.647	0.709
	appropriate-eval	131	0	4.58	0.626	3	5	-1.238	0.361	0.525	0.709
	aligns-approach	131	0	4.33	0.865	1	5	-1.627	0.361	3.756	0.709
	benefit-stakehldrs	131	0	4.44	0.700	3	5	-0.874	0.361	-0.434	0.709
	adds-value-efficy	131	0	4.33	0.837	2	5	-1.202	0.361	0.997	0.709
	have-used	131	0	2.86	1.317	1	5	-0.396	0.365	-1.253	0.717
AE	familiar-before	131	0	3.59	1.024	2	5	-0.505	0.350	-0.925	0.688
	learned-new	131	0	3.87	0.833	2	5	-0.711	0.350	0.355	0.688
	described-others	131	0	2.78	1.246	1	5	0.433	0.350	-1.066	0.688
	likely-to-use	131	0	3.96	0.842	2	5	-0.383	0.350	-0.488	0.688
	makes-sense	131	0	4.35	0.566	3	5	-0.128	0.350	-0.684	0.688
	appropriate-eval	131	0	4.30	0.628	3	5	-0.328	0.350	-0.596	0.688
	aligns-approach	131	0	4.22	0.728	3	5	-0.362	0.350	-1.004	0.688
	benefit-stakehldrs	131	0	4.33	0.634	3	5	-0.392	0.350	-0.616	0.688
	adds-value-efficy	131	0	4.20	0.749	3	5	-0.341	0.350	-1.116	0.688
	have-used	130	1	2.71	1.456	1	5	-0.024	0.354	-1.555	0.695

Note. Red indicates cut-offs were violated.